

**Development of Fluorescence-based Tools for
Characterization of Natural Organic Matter and
Development of Membrane Fouling Monitoring Strategies
for Drinking Water Treatment Systems**

by

Ramila Hishantha Peiris

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of

Doctor of Philosophy

in

Chemical Engineering

Waterloo, Ontario, Canada, 2010

© Ramila Hishantha Peiris 2010

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The objective of this research was to develop fluorescence-based tools that are suitable for performing rapid, accurate and direct characterization of natural organic matter (NOM) and colloidal/particulate substances present in natural water. Most available characterization methods are neither suitable for characterizing all the major NOM fractions such as protein-, humic acid-, fulvic acid- and polysaccharide-like substances as well as colloidal/particulate matter present in natural water nor are they suitable for rapid analyses. The individual and combined contributions of these NOM fractions and colloidal/particulate matter present in natural water contribute to membrane fouling, disinfection by-products formation and undesirable biological growth in drinking water treatment processes and distribution systems. The novel techniques developed in this research therefore, provide an avenue for improved understanding of these negative effects and proactive implementation of control and/or optimization strategies.

The fluorescence excitation-emission matrix (EEM) method was used for characterization of NOM and colloidal/particulate matter present in water. Unlike most NOM and colloidal/particulate matter characterization techniques, this method can provide fast and consistent analyses with high instrumental sensitivity. The feasibility of using this method for monitoring NOM at very low concentration levels was also demonstrated with an emphasis on optimizing the instrument parameters necessary to obtain reproducible fluorescence signals.

Partial least squares regression (PLS) was used to develop calibration models by correlating the fluorescence EEM intensities of water samples that contained surrogate NOM fractions with their corresponding dissolved organic carbon (DOC) concentrations. These fluorescence-based calibration models were found to be suitable for identifying/monitoring the extent of the relative changes that occur in different NOM fractions and the interactions between polysaccharide- and protein-like NOM in water treatment processes and distribution systems.

Principal component analysis (PCA) of fluorescence EEMs was identified as a viable tool for monitoring the performance of biological filtration as a pre-treatment step, as well as ultrafiltration (UF) and nanofiltration (NF) membrane systems. The principal components (PCs) extracted in this approach were related to the major membrane foulant groups such as humic substances (HS), protein-like and colloidal/particulate matter in natural water. The PC score plots generated using the fluorescence EEMs obtained after just one hour of UF or NF operation could be related to high fouling events likely caused by elevated levels of colloidal/particulate-like material in the biofilter effluents. This fluorescence EEM-based PCA approach was sensitive enough to be used at low organic carbon levels present in NF permeate and has potential as an early detection method to identify high fouling events, allowing appropriate operational countermeasures to be taken.

This fluorescence EEM-based PCA approach was also used to extract information relevant to reversible and irreversible membrane fouling behaviour in a bench-scale flat sheet cross flow UF process consisting of cycles of permeation and back-washing. PC score-based analysis

revealed that colloidal/particulate matter mostly contributed to reversible fouling, while HS and protein-like matter were largely responsible for irreversible fouling. This method therefore has potential for monitoring modes of membrane fouling in drinking water treatment applications.

The above approach was further improved by utilizing the evolution of the PC scores over the filtration time and relating these to membrane fouling by the use of PC scores' balanced-based differential equations. Using these equations the proposed fluorescence-based modeling approach was capable of forecasting UF fouling behaviours with good accuracy based solely on fluorescence data obtained at time = 15 min from the initiation of the filtration process. In addition, this approach was tested experimentally as a basis for optimization by modifying the UF back-washing times with the objective of minimizing energy consumption and maximizing water production. Preliminary optimization results demonstrated the potential of this approach to reduce power consumption by significant percentages. This approach was also useful for identifying the fouling components of the NOM that were contributing to reversible and irreversible membrane fouling.

Grand River water (Southwestern Ontario, Canada) was used as the natural water source for developing the techniques presented in this thesis. Future research focusing on testing these methods for monitoring of membrane fouling and treatment processes in large-scale drinking water treatment facilities that experience different sources of raw water would be useful for identifying the limitation of these techniques and areas for improvements.

Acknowledgements

First, I would like to express my heartfelt gratitude to my supervisors, Drs. Raymond Legge, Christine Moresoli and Hector Budman, for their continual support and guidance. Their directions have paved the way for a wonderful research experience which I will cherish for many years to come.

In addition, I would also like to thank Drs. William Anderson, Wayne Parker, Michel Perrier and Mark Pritzker for agreeing to be part of my PhD defence committee.

The work presented here is part of a larger project funded by the Canadian Water Network (CWN). The research funding rendered by CWN and the Natural Sciences and Engineering Research Council of Canada (NSERC) is very much appreciated. I am also grateful to NSERC for providing me with an NSERC Postgraduate scholarship. In addition, I thank the University of Waterloo for providing financial support through the President's Graduate Scholarship, International Doctoral Student Award and the University of Waterloo Special Graduate Scholarship.

I am very grateful to Dr. Peter Huck, Dr. Sigrid Peldszus and Cynthia Hallé of the Department of Civil and Environmental Engineering for the continued collaboration that helped me to have a broader scope for my research work. A special thank you to Cynthia for providing water samples from some of her own experimental work. The data from these samples were used to obtain the results presented in Chapters 7 and 8.

I am deeply thankful to Rand Elshereef who supported me by sharing his experience and knowledge on fluorescence spectroscopy and multivariate statistical methods. I would also like to thank all my lab colleagues for their part in making my research experience a pleasant and productive one.

Abubakar Masood and Erin Matheson are sincerely thanked for volunteering their time and assisting in some of the experimental work.

Finally, I am very grateful to my wife and my best friend Rochelle for her love, unwavering support and understanding during this research. Her encouragement, especially during difficult times, helped me to stay focused on my research. I also thank my daughter Ruwanthi for being a wonderful playmate when I needed a break! I am indebted to my parents, Ronald and Irene and my sister Heshani for their love, endurance, sacrifices and dedication in bringing me up to be the person I am. Last but not least, I thank my in-laws, Krishan, Ravindra, Premila and Ryan for their support and encouragement. I would not have been able to achieve this milestone without the help of these people who were dear and near to me.

Dedication

For my precious wife Rochelle Peiris who offered me her unconditional love and continuous encouragement.

Table of Contents

List of Figures	xviii
List of Tables	xxiii
Nomenclature	xxv
CHAPTER 1 Research Goals and Outline.....	1
1.1 Research Motivation	1
1.2 Research Objectives	3
1.3 Thesis Structure.....	4
CHAPTER 2 Theoretical Background.....	8
2.1 Application of Membranes in Drinking Water Treatment	8
2.2 Membrane Fouling in Drinking Water Applications	9
2.2.1 Membrane Fouling Control Strategies.....	10
2.2.2 Major Membrane Foulants and Factors Affecting Fouling	11
2.2.3 Theoretical Basis of Membrane Fouling.....	12
2.3 Membrane Fouling Modeling.....	13
2.3.1 Mechanistic Membrane Fouling Modeling Approaches.....	13
2.3.2 Non-mechanistic Membrane Fouling Modeling Approaches.....	15
2.4 Optimization of Membrane Filtration Processes for Drinking Water Treatment	15
2.5 Natural Organic Matter	16
2.5.1 Natural Organic Matter as a Precursor for Disinfected By-product Formation.....	17
2.6 Significance of NOM Characterization.....	18
2.7 Characterization of Natural Organic Matter Foulants.....	19
2.8 Principles of Fluorescence Spectroscopy	20

2.8.1	Fluorophores	21
2.8.2	Theory of Fluorescence.....	21
2.9	Characterization of NOM by Fluorescence Spectroscopy	24
2.9.1	Characterization of Particulate/colloidal Matter using Rayleigh Scattering	24
2.9.2	Effect of the Solution Properties, the Season and the Source of Water on Fluorescence Spectral Characteristics	25
2.10	Techniques that Characterize the Properties of Fluorescence Spectra.....	26
2.11	Multivariate Data Analysis of Fluorescence Excitation-emission Matrices	27
2.11.1	Partial Least Squares (PLS) Regression	28
2.11.1.1	Significance of PLS Compared to Other Classical Methods of Statistics.....	29
2.11.1.2	Comparison Between PLS and Other Multivariate Regression Methods	30
2.11.1.3	PLS Regression Methodology	31
2.11.1.3.1	Pre-processing of Data	31
2.11.1.3.2	Theory	32
2.11.1.3.3	Finding the Optimum Number of PLS Components (or LVs).....	35
2.11.2	Principal Component Analysis	36
2.11.2.1	PCA Analysis Methodology	36
2.11.2.1.1	Pre-processing of Data	36
2.11.2.1.2	Theory	36
2.11.2.1.3	Selecting Statistically Significant PCs	37
CHAPTER 3 Acquiring Reproducible Fluorescence Spectra of Dissolved Organic Matter at Very Low Concentrations		39
3.1	Introduction	40
3.2	Methods.....	42
3.2.1	Water Samples and Preparation	42

3.2.2	Fluorescence Analysis	43
3.3	Results and Discussion.....	44
3.3.1	PMT Voltage.....	47
3.3.2	Scanning Rate	48
3.3.3	Emission and Excitation Slit Widths	49
3.4	Conclusions	54
CHAPTER 4 Identification of Humic Acid-like and Fulvic Acid-like Natural Organic Matter in River Water using Fluorescence Spectroscopy		56
4.1	Introduction	57
4.2	Methods.....	59
4.2.1	Sample Preparation	59
4.2.2	Fluorescence Analysis	62
4.3	Results and Discussion.....	63
4.3.1	Typical Spectral Features of GRW	63
4.3.2	Fluorescence EEMs of FA-like and HA-like NOM in GRW	64
4.3.3	Significance of this Approach and Potential for Future Research.....	67
4.4	Conclusions	68
CHAPTER 5 Development of a Fluorescence-based Soft Sensor Approach to Differentiate and Quantify Natural Organic Matter in Water.....		69
5.1	Introduction	70
5.2	Materials and Methods	74
5.2.1	Sample Preparation.	74
5.2.2	Calibration Sample Sets.....	75
5.2.3	Validation Sample Sets.....	75
5.2.4	Fluorescence Analysis	77
5.2.5	Soft-sensor Development.....	77

5.3	Results and Discussion.....	79
5.3.1	Calibration Model Development.....	80
5.3.2	Model Predictions within Calibration Range.....	81
5.3.3	Model Predictions outside Calibration Range	83
5.3.4	Physical Relevance of Latent Variables	84
5.3.5	Significance of the FSS Approach Compared to Other Fluorescence-based Methods	87
5.4	Conclusions	87
5.5	Supplementary Information.....	89
5.5.1	The Non-linear Behaviour of the Predictions at High AHA Concentrations	89
5.5.2	Quantification of AHA and GRW using a Calibration Model Based on FRI Approach	90
5.5.3	Significance of Fluorescence-based Soft Sensor Approach Compared to Analyzing Only Major Ex/Em Peak Positions (Peak-Picking Approach).....	93
CHAPTER 6 Fluorescence-based Rapid Identification of Different Natural Organic Matter Fractions for Drinking Water Treatment Applications		94
6.1	Introduction	95
6.2	Methods.....	99
6.2.1	Experimental Approach and Rationale	99
6.2.2	Sample Preparation.	100
6.2.3	DOC and LC-OD analyses.....	101
6.2.4	Fluorescence Analysis	101
6.2.5	Calibration Set	102
6.2.6	Validation Set.....	103
6.2.7	Analysis of Variance (ANOVA) of Fluorescence EEMs	104
6.2.8	Model Development and Validation.....	104

6.3	Results and Discussion.....	106
6.3.1	LC-OCD Characterization of Major Surrogate NOM Fractions	106
6.3.2	Fluorescence EEMs of Surrogate NOM	107
6.3.3	Model Predictions	110
6.3.4	Significance of the Approach for Identification of Major NOM Fractions.....	113
6.3.5	Characterizing the Interactions between Protein-like and Polysaccharide-like NOM Fractions.....	114
6.3.6	Potential for Process Monitoring and Other Applications	115
6.4	Conclusions	116
6.5	Supplementary Information.....	118
6.5.1	Optimum Scanning Path Approach for Fluorescence Signal Acquisition.....	121
CHAPTER 7 Assessing Nanofiltration Fouling in Drinking Water Treatment Using Fluorescence Fingerprinting and LC-OCD Analyses		124
7.1	Introduction	125
7.2	Methods.....	127
7.2.1	Water Samples and NF Set-up.....	127
7.2.2	Analytical Methods.....	129
7.2.3	Fluorescence Analysis	130
7.3	Results and Discussion.....	130
7.4	Conclusions	136
CHAPTER 8 Identifying Fouling Events in a Membrane-based Drinking Water Treatment Process using Principal Component Analysis of Fluorescence Excitation-Emission Matrices.....		138
8.1	Introduction	139
8.2	Materials and Methods	141
8.2.1	Feed Water and Pre-treatment	141

8.2.2	Pilot-scale Membrane Filtration Set-up	143
8.2.2.1	Ultrafiltration Membrane	143
8.2.2.2	Nanofiltration Membrane	144
8.2.3	Fluorescence Analysis	145
8.2.4	Fluorescence Data Pre-treatment and PC Analysis	146
8.3	Results and Discussion	148
8.3.1	Typical Spectral Features in the Fluorescence EEM of GRW	148
8.3.2	PCs that Summarize the Total Variance Captured in the Fluorescence EEMs	149
8.3.3	Physical Significance of the PCs Generated by PCA	150
8.3.4	Performance of the Pre-treatment and UF/NF Stages as Summarized by the Score Plots of PCs and Potential as Performance Monitoring Tool.	153
8.3.5	Identifying High Fouling Events by PCA of Fluorescence EEMs	161
8.3.6	Potential for Process Monitoring and Intervention	163
8.4	Conclusions	164
8.5	Supplementary Information.....	166
CHAPTER 9 Understanding Fouling Behaviour of Ultrafiltration Membrane Processes and Natural Water using Principal Component Analysis of Fluorescence Excitation-Emission Matrices.....		170
9.1	Introduction	171
9.1.1	Natural Water and Membrane Foulants	172
9.1.2	Fluorescence-based Fouling Monitoring Approach.....	173
9.2	Materials and Methods	176
9.2.1	Feed Water and Pre-treatment	176
9.2.2	Bench-scale Membrane Ultrafiltration Set-up	176
9.2.3	Fluorescence Analysis	178
9.2.4	Other Analytical Methods	179

9.2.5	Extraction of Membrane Foulants	179
9.2.6	Fluorescence Data Pre-treatment and PCA.....	180
9.2.7	PC Score-based Evaluation of Membrane Fouling.....	182
9.3	Results and Discussion.....	184
9.3.1	Typical Fluorescence EEM Features of GRW.....	184
9.3.2	PCA of Fluorescence EEMs	185
9.3.3	PCA of Fluorescence Data Identifies Foulant Components	186
9.3.4	Different Fouling Behaviour Captured by PC Scores.....	189
9.3.5	Fluorescence EEMs of Accumulated Membrane Fouling Components	194
9.3.5.1	Colloidal/particulate Matter Fouling	195
9.3.5.2	Fouling by HS-like Components	196
9.3.5.3	Fouling by Protein-like Components.....	197
9.3.5.4	Different Membrane Rejection Characteristics	198
9.3.6	Potential for Process Monitoring and Future Applications.....	200
9.4	Conclusions	200
CHAPTER 10 Fluorescence-based Fouling Prediction and Optimization of a Membrane Filtration Process for Drinking Water Treatment.....		203
10.1	Introduction	204
10.1.1	Methods Available for Membrane Fouling Modeling	205
10.1.2	Proposed Fouling Modeling and Optimization Approach.....	206
10.2	Materials and Methods	208
10.2.1	Bench-scale Membrane Filtration Set-up	208
10.2.2	Fluorescence Analysis	210
10.2.3	Other Analytical Methods.....	211
10.2.4	Fluorescence Data Pre-treatment and PCA.....	211

10.3	Theoretical Basis	212
10.3.1	PC-based Modeling of Membrane Fouling.....	213
10.3.2	Model Calibration and Validation	216
10.3.3	Generation of Model Predictions	217
10.3.4	Optimization of the UF Process	218
10.4	Results and Discussion.....	220
10.4.1	Typical Fluorescence Spectral Features of Grand River Water.....	220
10.4.2	PCA of Fluorescence Data.....	220
10.4.3	Model Predictions	222
10.4.4	Optimization of UF for Drinking Water Treatment.....	223
10.4.5	Role of Membrane Foulant Components in Reversible and Irreversible Fouling	225
10.5	Conclusions	228
10.6	Supplementary Information.....	229
10.6.1	Typical Fluorescence Spectral Features of Grand River Water.....	229
10.6.2	Loading Plots for 60 kDa and 20 kDa UF Membrane Experiment	230
10.6.3	Model Parameters for Fluorescence-based Fouling Modeling Approach	232
CHAPTER 11	Conclusions & Recommendations.....	233
11.1	Acquiring Reproducible Fluorescence Signals for Very Low NOM Concentration Levels.....	234
11.2	Fluorescence-based Qualitative Characterization of NOM	234
11.3	Fluorescence-based Calibration Models for Characterization of NOM.....	235
11.4	Process Monitoring in Drinking Water Treatment Systems using Fluorescence EEM Measurements	236
11.5	Understanding Membrane Fouling Behaviour	237
11.6	Membrane Fouling Modeling and Optimization.....	238

11.7 Future Research Work.....	239
REFERENCES	242
APPENDICES	
APPENDIX 1 Rationale for the Selection of PMMA Disposable Cuvettes.....	262
APPENDIX 2 Selected Photographs from UF Experiments	266
APPENDIX 3 Contact Angle Measurements of 20 kDa and 60 kDa UF Membranes at Different Stages.....	270
APPENDIX 4 Typical LC-OCD Chromatographs for the Identification of Ultrafiltration Membrane Foulants	271

List of Figures

Figure 2-1 Schematic representation of the dead-end and cross flow filtration operations.....	8
Figure 2-2 Schematic representation of membrane fouling.....	10
Figure 2-3 A simplified Jablonski diagram that illustrates the processes that occur between absorption of light and generation of fluorescence signals.	23
Figure 2-4 Schematic representation of the derivation of PLS components and the corresponding scores in multivariate explanatory and multivariate response spaces.....	33
Figure 2-5 The matrix relationship in PLS regression. The score, weight and loading matrices are derived during the development of the PLS regression model.	34
Figure 3-1 Fluorescence EEMs of GRW obtained with (a) cuvettes and (b) fibre optic probe. Fluorescence EEMs of GRW diluted up to 10% of the original strength obtained with (c) cuvettes and (d) fibre optic probe. All measurements were performed at scanning rate of 600 nm/min, PMT voltage = 800 V and excitation and emission slit widths of 5 nm each.	45
Figure 3-2 Trend lines of excitation and emission spectra of GRW, diluted up to 10% of its original strength, taken at emission = 415 nm and excitation = 330 nm respectively using the fibre optic probe. A scanning rate of 600 nm/min, PMT voltage = 800 V and excitation and emission slit widths of 10 nm each were used.	52
Figure 3-3 Emission spectra of GRW at excitation = 330 nm with different excitation and emission slit width combinations obtained using the fibre optic probe at a scanning rate of 600 nm/min and PMT voltage = 800 V.	53
Figure 4-1 Procedure for the fractionation of HA- and FA-like NOM fractions from GRW.....	61
Figure 4-2 Fluorescence EEM for GRW (left) and 3D view of the same EEM (right).....	63
Figure 4-3 Fluorescence EEMs of (a) FA-like NOM – obtained at pH ~ 8.2; (b) HA-like NOM extract at pH < 1 and (c) HA-like NOM extract at pH ~ 8.2.....	65
Figure 4-4 Fluorescence EEM of Aldrich humic acid obtained at pH = 8.2.	66

Figure 5-1 Fluorescence EEM of (a) AHA (b) GRW and (c) a mixture of AHA and GRW.....	80
Figure 5-2 Predictions of the quantity of AHA mixed into GRW, diluted up to (a) 10%, (b) 25%, (c) 50% and (d) 75% of its original strength.	82
Figure 5-3 Predictions of the quantity of GRW in different mixtures of GRW and AHA.....	83
Figure 5-4 Weight loading values plotted at their corresponding excitation/emission wavelengths; (a) weight loading matrix on the first latent variable (LV1), (b) 3D representation of weight loadings on LV1, (c) weight loading matrix on the second latent variable (LV2) and (d) 3D representation of weight loadings on LV2.....	85
Figure 6-1 Experimental cube design (ECD) used in the sample preparation for calibration and validation sample sets.	103
Figure 6-2 LC-OCD chromatograms of (a) Grand River water and (b) surrogate NOM fractions used in the development of the fluorescence-based calibration model.	106
Figure 6-3 Fluorescence EEMs of (a) GRW and GRW spiked with (b) AHA, (c) AA and (d) AL.....	107
Figure 6-4 Comparisons between the model predictions and the individual measured DOC values of (a) AHA, (b) AL (c) AA and (d) GRW in water samples in the validation sample set. This validation sample set contains different concentration combinations of AHA, AL, AA and GRW.....	112
Figure 6-5 Comparisons between the model predictions and the estimated values for interaction between AL and AA based on measured DOC.	115
Figure 7-1 Schematic of the nanofiltration experimental set-up.....	129
Figure 7-2 Fluorescence EEMs of (a) GRW (RW) compared with (b) corresponding LC-OCD spectra.	131
Figure 7-3 Fluorescence EEMs for (a) GRW filtered through the roughing filter (RF) and (b) biofilter effluent.....	131
Figure 7-4 Fluorescence EEMs of the membrane feed (b) and the permeates of the NF membranes; XN45 (c) and TS80 (d) and the corresponding LC-OCD spectra (a).....	134

Figure 8-1 Pilot scale membrane filtration experimental set-up.....	143
Figure 8-2 Typical fluorescence features seen in the (a) fluorescence EEM for GRW and (b) 3D view of the same EEM.	149
Figure 8-3 Loading plots of (a) PC – 1 - related to the humic content, (b) PC – 2 - related to the particulate/colloidal content and (c) PC – 3 - related to the protein content in water.....	151
Figure 8-4 Score plots: (a) PC – 2 vs. PC – 1 and (b) PC – 3 vs. PC – 2. PC scores are grouped and named based on the sampling locations: RW, RF, B1, B2 and UFp (Figure 8-1).....	155
Figure 8-5 Score plot: (a) PC – 2 vs. PC – 1 and (b) PC – 3 vs. PC – 2. PC scores are grouped and named based on the sampling locations: RW, RF, B1, B2, NF_C, NF_tank and NFp.	160
Figure 9-1 Schematic of the bench-scale ultrafiltration cross flow set-up.	177
Figure 9-2 Typical fluorescence features seen in the (a) fluorescence EEM for GRW and (b) 3D view of the same EEM.	185
Figure 9-3 3D illustrations of the loading matrices of (a) PC – 1, (b) PC – 2, (c) PC – 3 and (d) PC – 4 generated by the PCA of 60 kDa data.	187
Figure 9-4 Normalized UF permeate flux profiles for (a) 60 kDa and (b) 20 kDa membranes with different feed water quality and fouling potential.....	190
Figure 9-5 Plots of normalized PC score difference of retentate and permeate ($\Delta S_{j, \text{Norm}}$) vs. filtration time for 60 kDa (left: a, b, c and d) and 20 kDa (right: e, f, and g) membranes.....	191
Figure 9-6 Fluorescence EEMs of the loosely-attached foulants (above) and ethanol extracted foulants (below) for 60 kDa (a and b) and 20 kDa (c and d) membranes.....	197
Figure 10-1 Bench-scale ultrafiltration cross flow set-up.....	209
Figure 10-2 Model predictions (lines) and experimentally measured (symbols) normalized permeate water flux for selected (a) 60 kDa and (b) 20 kDa UF experiments of low, medium and high membrane fouling situations.....	222
Figure 10-3 Model predictions (lines) and experimentally measured (symbols) normalized permeate water flux obtained for (a) 60 kDa and (b) 20 kDa UF operations with normal back-washing (BW) times (every hour) and optimized back-washing times.	223

Figure 10-4 The contribution of (a) humic substances (HS) – like (PC – 1), (b and d) colloidal/particulate (PC – 2 and PC – 4) and (c) protein-like (PC – 3) matter on membrane resistance as calculated in terms of the accumulation of individual foulant components in/on the membranes for selected 60 kDa UF experiments of low, medium and high membrane fouling situations.....	226
---	-----

Supplementary Figures

Figure S - 5-1 An example of non-linear behaviour manifested in the predictions of AHA: (a) prediction of AHA concentration represented as function of calculated AHA concentration and (b) residual plot of model predictions.	89
Figure S - 5-2 Two regions of the fluorescence EEM used in the FRI analysis. (a) EEM of GRW and (b) weight loading matrix of LV2.....	91
Figure S - 5-3 Comparison of main peak intensities of (a) AHA (Ex/Em: 270nm /460 nm) and (b) GRW (Ex/Em: 320nm /415 nm) at different concentration levels for randomly selected samples in the calibration set presented in this chapter.....	93
Figure S - 6-1 Steps involved in generating the fluorescence-based calibration model.	118
Figure S - 6-2 Fluorescence EEM of alginic acid (20 mg/L), which demonstrates that alginic acid does not fluoresce.....	119
Figure S - 6-3 Sensitivity spectra of the PLS-based calibration model that illustrate areas in the fluorescence EEMs which are relevant for predicting the AHA, AL, AA and GRW concentrations.	121
Figure S - 6-4 Optimum scanning path that was used to minimize the fluorescence acquisition time.....	122
Figure S - 8-1 Steps involved in generating data rows of intensity values for principal component analysis (PCA).	168
Figure S - 8-2 A comparison of the JCR obtained by (a,b) PCA of the full fluorescence EEM intensity data and (c,d) PCA of only the main peak positions of the fluorescence EEMs such as peaks (α), (β), (δ) and the Rayleigh scattering peaks recorded during the UF experiment.....	169
Figure S - 10-1 Typical fluorescence features seen in the (a) fluorescence EEM for GRW and (b) 3D view of the same EEM.	229

Figure S - 10-2 3D illustrations of the loading matrices of (a) PC – 1, (b) PC – 2, (c) PC – 3 and (d) PC – 4 generated by the PCA of 60 kDa data.	230
Figure S - 10-3 3D illustrations of the loading matrices of PC -1, PC-2 and PC-3 generated by the PCA of 20 kDa data.	231

Appendices Figures

Figure A - 1 Fluorescence EEMs of GRW taken with (a) Quartz cuvettes (above) and (b) PMMA cuvettes (below).	263
Figure A - 2 Fluorescence EEMs of Aldrich humic acid (4 mg/L) taken with (a) Quartz cuvettes (above) and (b) PMMA cuvettes (below).	264
Figure A - 3 Comparison of the fluorescence EEM cross sections of (a) AHA (taken at emission = 475 nm) and (b) GRW (taken at emission = 415 nm) obtained with PMMA and Quartz cuvettes.	265
Figure A - 4 Photographs of bench-scale cross flow UF set-up. (a) A view of the UF membrane cell holder including other instrumentation and (b) feed tank connected to the diaphragm pump.	266
Figure A - 5 (a) A close-up view of the UF instrumentation panel and (b) a computer display of the LabView–based interface used for data logging and monitoring purposes.	267
Figure A - 6 Photographs showing (a) sample collection for fluorescence analysis during permeation (b) back-washing of membranes assisted by compressed nitrogen gas.	268
Figure A - 7 Photographs showing membrane surfaces of (a and d) fouled, (b and e) physically cleaned and (c and f) chemically cleaned (i.e. using ethanol) membranes for 60 kDa (left) and 20 kDa (right) UF membranes.	269
Figure A - 8 LC-OCD chromatographs of feed water (filtered GRW using a 200 micron filter), permeate and loosely attached foulant (LAF 1) material obtained for a 20 kDa UF experiment conducted using GRW obtained on November 23, 2009. LAF 2 was collected from a 60 kDa UF experiment conducted using GRW obtained on October 25, 2009.	271

List of Tables

Table 3-1 Measurement error observed with the fibre optic probe vs. cuvettes for different dilutions of GRW at PMT = 800 V, scanning rate of 600 nm/min and excitation and emission slit widths of 5 nm each.	47
Table 3-2 Measurement error observed with the fibre optic probe and cuvettes for two different dilution levels of GRW at PMT = 1000V, scanning rate of 600 nm/min and excitation and emission slit widths of 5 nm each.	48
Table 3-3 Measurement error observed when signal was acquired using a fibre optic probe for two different dilution levels of GRW at PMT = 800 V, scanning rate of 120 nm/min and excitation and emission slit widths of 5 nm each.	49
Table 3-4 Signal noise at different slit width combinations, obtained with the fibre optic probe and cuvettes on GRW (diluted up to 10%) – calculated at the EEM Peak at Ex: 330 nm; Em: 415 nm.	50
Table 5-1 Concentrations of AHA and GRW in the calibration and validation sample sets.	76
Table 6-1 Results of the ANOVA that illustrates the effect of NOM fraction surrogates on fluorescence EEM intensities.	110
Table 7-1 Characteristics of Grand River water for the period August to September 2007.	127
Table 7-2 Fluorescence intensities at reported peak positions for HS and protein-like substances.	132
Table 8-1 Grand River water quality parameters from August 2007 – August 2008.	142
Table 8-2 Variance captured by the first three principal components.	150
Table 8-3 Comparison of typical TOC, DOC and turbidity values under normal filtration conditions and the values recorded under high fouling events.	157
Table 9-1 UF feed water quality parameters from March – October, 2009.....	176
Table 9-2 Variance captured in the PCA of X60 and X20 matrices.	186
Table 9-3 Fluorescence EEM peak intensities of the ethanol extracted foulants for 60 kDa and 20 kDa membranes.	199

Table 10-1 Variance captured in the PCA of the X60 and X20 matrices.	221
--	-----

Supplementary Tables

Table S - 5-1 Comparison of model prediction error in terms of the total variation in DOC concentration for GRW and AHA between fluorescence-based soft sensor and FRI approaches.	92
Table S - 6-1 Results of the ANOVA that illustrates the effect of surrogate NOM fractions on fluorescence EEM intensities.	120
Table S - 8-1 Details of the ultrafiltration experiments.	166
Table S - 8-2 Details of the nanofiltration experiments.	167
Table S - 10-1 Parameters used in the modeling of 60kDa and 20kDa UF processes.....	232

Appendices Tables

Table A - 1 Contact angle measurements of 20kDa and 60 kDa UF membranes at different stages.....	270
---	-----

Nomenclature

^{13}C	- Isotope of carbon with 7 neutrons and 6 protons
3D	- Three dimensional
A	- Effective membrane surface area (m^2)
AA	- Alginic acid
AHA	- Aldrich humic acid
AL	- α -lactalbumin
ANOVA	- Analysis of variance
a.u.	- Arbitrary units
BF	- Biofilter effluent
BW	- Back-washing
C	- Matrix containing \mathbf{c}_i vectors
CCA	- Correspondence analysis
\mathbf{c}_i	- Vector containing loadings of i^{th} Latent variable in Y space
CI	- Confidence interval
D	- Effective diffusivity (m/s)
DBP	- Disinfection by-products
DI	- De-ionized
DOC	- Dissolved organic carbon (mg/L)
DOM	- Dissolved organic matter
E	- Residual matrix resulting from principal component analysis
ECD	- Experimental cube design

EEM	- Excitation-emission matrix
<i>eff</i>	- Efficiency at which membrane foulants are removed during back-washing.
Em	- Emission wavelength (nm)
E _o	- Residual matrix resulting from partial least squares regression for explanatory variable space
Ex	- Excitation wavelength (nm)
FA	- Fulvic acid
FM	- Flow meter
F _o	- Residual matrix resulting from partial least squares regression for response variable space
FORS	- First order Rayleigh scattering
FRI	- Fluorescence regional integration
FSS	- Fluorescence-based soft sensor
FTIR	- Fourier transform infrared
GC	- Gas chromatography
GRW	- Grand River water
H	- Ultrafiltration experiments identified as a high fouling event
HA	- Humic acid
HPSEC	- High performance size exclusion chromatography
HS	- Humic substances
IR	- infrared
JCR	- Joint confidence region

J_o	- Pure water flux across the membrane (L/min.m ²)
J_t	- Permeate flux at time = t (L/min.m ²)
J_t/J_o	- Normalized membrane flux
k	- Parameter that specifies the active portion of V_M
L	- Ultrafiltration experiments identified as a low fouling event
LC-OCD	- Liquid chromatography - organic carbon detection
LMH	- Litre per meter cubed per hour
LV	- Latent Variable
\dot{m}	- Volume flow rate (m ³ s ⁻¹)
M	- Ultrafiltration experiments identified as a medium fouling event
MF	- Microfiltration
MLR	- Multiple linear regression
MPE	- Maximum prediction error
MS	- Mass spectrometry
MW	- Molecular weight (g/mol)
MWCO	- Molecular weight cut-off (kDa)
MW _n	- Number averaged molecular weight (g/mol)
MW _w	- Weight averaged molecular weight (g/mol)
n	- Number of both retentate and permeate samples
N	- Number of statistically significant principal components
NF	- Nanofiltration
NMR	- Nuclear magnetic resonance
NOM	- Natural organic matter

NTU	- Nephelometric Turbidity Units
ODE	- Ordinary differential equation
OF	- Objective function
PARAFAC	- Parallel factor
PC	- Principal component
PCA	- Principal component analysis
PCR	- Principal component regression
PG	- Pressure gauge
\mathbf{p}_i	- Vector containing loading values of i^{th} principal component
PLS	- Partial least squares
PMMA	- Polymethylmethacrylate
PMT	- Photomultiplier tube
PT	- Pressure transducer
PVC	- Polyvinyl chloride
q	- Parameter describing the decay of efficiency in back-washing
\mathbf{Q}	- Matrix containing \mathbf{q}_i vectors
Q^2	- Goodness of prediction
\mathbf{q}_i	- Vector containing loadings of i^{th} Latent variable in X space
R	- Membrane resistance
R^2	- Goodness of fit
RDA	- Redundancy analysis
RF	- Effluent of roughing filter
RO	- Reverse osmosis

RS	- Rayleigh light scattering regions
RW	- Untreated (raw) GRW
\mathbf{s}_i	- Vector containing scores of i^{th} principal component
SORS	- Second order Rayleigh scattering
SUVA	- Specific (UV) absorbance (L/mg-M)
T	- Matrix containing \mathbf{t}_i vectors
t	- Time (s or min)
\mathbf{t}_i	- Vector containing scores of i^{th} Latent variable in X space
t_d	- Time at which the first set of fluorescence EEMs of the retentate and permeate for UF operation were collected.
t_F	- Total filtration period (s or min)
tw	- Back-washing period (s)
TMP	- Trans-membrane pressure (psi – pressure per square inch / kPa)
TOC	- Total organic carbon (mg/L)
U	- Matrix containing \mathbf{u}_i vectors
UF	- Ultrafiltration
\mathbf{u}_i	- Vector containing scores of i^{th} Latent variable in Y space
UV	- Ultraviolet
V_M	- Volume of the solution occupied by the membrane (m^3)
w	- Binary variable that models permeation through the membrane (w = 0) or back-washing (w = 1).
W	- Weight matrix that explains the correlation between X and U

X	- Matrix containing fluorescence EEM data (explanatory/predictor variables)
X20	- Matrix containing fluorescence EEM data for 20 kDa UF membrane experiments
X60	- Matrix containing fluorescence EEM data for 60 kDa UF membrane experiments
Xc	- Fluorescence EEM data set for calibration
Xv	- Fluorescence EEM data set for validation
Y	- Matrix of response variables
Yc	- Response data set for calibration
z	- Parameter that serves for updating D_i

Greek letters

α	- Fluorescence EEM peak representing humic substances
α'	- Loading peak related to humic substances
β	- Fluorescence EEM shoulder to peak (α) representing humic substances
β	- Parameter that relates PC score of a foulant to membrane resistance
β'	- Loading peak (shoulder to peak α') related to humic substances
ΔP	- Tran-membrane pressure (psi – pressure per square inch / kPa)
ΔS_j	- Difference between the score of retentate and permeate of jth principal component; $\Delta S_j = (S_{j,R} - S_{j,P})$

$\Delta s_{j, Norm}$	- Normalized ΔS_j with respect to the PC scores of the initial feed water (s_{j, F_0})
Δt	- A given time interval/period (min)
δ	- Fluorescence EEM peak representing protein-like substances
δ'	- Loading peak related to protein-like substances
δ^1	- Fluorescence EEM peak related to protein-like matter in ethanol-based foulant for 60 kDa membrane
δ^2	- Fluorescence EEM peak related to protein-like matter in ethanol-based foulant for 20 kDa membrane
ϵ_{280}	- molar extinction coefficient
μ	- Water viscosity (Nsm^{-2})

Subscripts

0	- Initial (time= 0 s)
c	- Calibration
Coll./partic.	- Colloidal/particulate
F_0	- Initial feed water
i	- Sample number in X matrix containing fluorescence EEMs
<i>init</i>	- Initial
<i>inter</i>	- Interaction
j	- Statistically significant principal component number
M	- Membrane
Norm	- Normalized

P	- Permeate
R	- Retentate
t	- Time
v	- Validation
wash	- Membrane back-washing

CHAPTER 1

Research Goals and Outline

1.1 Research Motivation

Membranes are increasingly used for the production of drinking water as effective physical barriers for particulate matter and for the improved removal of pathogenic organisms, salinity as well as a wide range of contaminants including pesticides and metals. Membrane fouling is a major obstacle for maintaining efficient performance levels of membrane-based drinking water treatment operations. Fouling contributes to increased operational costs associated with permeate flux decline and/or increased trans-membrane pressure (TMP) requirements. Frequent chemical cleaning of fouled membranes also deteriorates membrane performance, leading to shortened service life. Controlling membrane fouling is therefore of paramount importance to ensure the sustainable operation of membrane-based drinking water treatment processes.

Natural organic matter (NOM) and colloidal/particulate substances present in natural water are major membrane foulants in drinking water treatment applications (Jermann *et al.*, 2007). NOM that may pass through membranes also contributes to disinfection by-products (DBP) formation and increases the need for maintaining an effective disinfectant residual concentration during treatment and in distribution systems. But the complex nature and variability of NOM and colloidal/particulate matter complicates the development of fouling mitigation strategies. Therefore, characterization of NOM and colloidal/particulate matter

present in natural water is essential for understanding and developing strategies for controlling membrane fouling as well as addressing consumer-related water quality issues.

Most of the currently available NOM characterization methods are not suitable for performing rapid, accurate and direct characterizations (Her *et al.*, 2003; Croue, 2004) which are important when developing monitoring tools. Fluorescence spectroscopy has been identified as a promising method for NOM characterization and has good potential for on-line or rapid off-line monitoring with high instrumental sensitivity (Coble *et al.*, 1990). However, implementation of this technique as an effective NOM and colloidal/particulate matter monitoring tool for drinking water treatment operations has not been established. This can be attributed to the lack of scientific understanding of the application of fluorescence data analysis methods (Henderson *et al.*, 2009) for extracting information related to NOM and colloidal/particulate matter present in natural water.

In addition, development of fouling control strategies that are built on fluorescence-based characterization of membrane foulants could be of significant value for the drinking water industry that uses membrane-based treatment operations. Such strategies would also allow better understanding of physicochemical phenomena involved in membrane fouling from a fundamental perspective – an area that has been poorly understood thus far and has significant bearing on the development of effective fouling modeling approaches (Lee *et al.*, 2009).

1.2 Research Objectives

The main objective of this research was to develop fluorescence-based tools for rapid characterization of NOM and colloidal/particulate matter present in drinking water treatment processes. These characterization approaches were also expected to serve as a basis for developing membrane fouling monitoring tools for drinking water treatment systems. The following research objectives were identified to accomplish this global objective:

1. Obtain reproducible fluorescence spectra especially for very low NOM concentrations comparable to those found in permeates of nanofiltration (NF) and ultrafiltration (UF) membrane systems.
2. Develop fluorescence-based calibration models which could be used for rapidly identifying the extent of the relative changes that occur in different NOM fractions in drinking water treatment processes and distribution systems.
3. Investigate the different fouling behaviours exhibited by different membrane types using fluorescence spectroscopy.
4. Develop a fluorescence-based process monitoring approach to rapidly identify membrane fouling events (i.e. high fouling events), resulting from the changes in the performance levels of different pre-treatment stages, in drinking water treatment operations.
5. Investigate reversible and irreversible fouling behaviour in UF-based membrane treatment of drinking water using fluorescence-based measurements.
6. Develop fluorescence-based modeling tools for fouling prediction and optimization of UF membrane processes for drinking water treatment.

1.3 Thesis Structure

This thesis includes eleven chapters, a list of references and appendices. Chapter 1 outlines the research motivations, research objectives and the thesis structure and is followed by Chapter 2 which consists of background information relating to the topics that are covered in this thesis. Chapters 3 – 10 are presented in manuscript format, each starting with a short introduction. As a result of this format, there will be a repetitive nature to some of the information found in the introduction and methods sections of these chapters. Chapters 3 – 6 presents results related to the fluorescence-based characterization of NOM. Approaches developed for monitoring, understanding and modeling of NOM and colloidal/particulate matter in relation to membrane fouling in drinking applications is presented in Chapters 7 – 10. Overall conclusions and recommendations are summarized in Chapter 11. The contents of each of the chapters are as follows:

Chapter 3 – Direct application of fluorescence spectroscopy for characterization of very low concentration levels of NOM, such as the levels found in nanofiltration permeates, is difficult although this capability is essential for successful implementation of fluorescence-based NOM characterization and fouling monitoring strategies in drinking water applications. This chapter describes an approach for obtaining reproducible fluorescence spectra at very low dissolved NOM concentration levels. The importance of optimizing different fluorescence spectrofluorometer parameters is discussed. The optimum parameters identified in this study were used throughout this thesis

for characterization of NOM. This chapter has been published in *Water Science and Technology* (Peiris *et al.*, 2009).

Chapter 4 – Identifying the extent of major NOM fractions present in natural water is important to assess disinfection by-products formation and fouling potential during drinking water treatment applications. This chapter focuses on the qualitative characterization of two main NOM fractions: humic acid (HA) - like and fulvic acid (FA)-like matter, extracted from natural river water, using fluorescence spectroscopy. The material in this chapter has been submitted to *IWA World Water Congress, 2010* to be held in Montréal, Canada from September 19 - 24. This material is also being considered for publication in *Water Science and Technology* or *Water Science and Technology: Water Supply* (Submission date: September 23, 2009).

Chapter 5 – Most reported NOM characterization techniques, including fluorescence-based methods, have focused on qualitative characterization of NOM but not on quantification of the different NOM fractions in water. This is due to the complex heterogeneity related to different NOM fractions in natural water and the difficulty associated with the isolation of different NOM fractions. In this context, this chapter explains the development of a fluorescence-based calibration modeling methodology as a method for direct and rapid quantification and differentiation of two major NOM fractions (i.e. FA- and

HA-like NOM) present in water. This chapter has been prepared for submission.

Chapter 6 – This chapter focuses on improving the approach presented in Chapter 5 for developing a fluorescence-based calibration model capable of rapidly identifying all the major NOM foulant fractions such as HA-, FA-, protein- and polysaccharide-like material present in drinking water treatment processes. This chapter has been prepared for submission to *Water Research*.

Chapter 7 – This chapter demonstrates how fluorescence spectroscopy can be used to qualitatively assess NF membrane fouling as applied in drinking water treatment. In particular, the differences seen in the fluorescence spectra of permeates of two different NF membrane types corroborated the observed fouling behaviour. This chapter has been published in the journal *Water Science and Technology: Water Supply* (Peiris *et al.*, 2008).

Chapter 8 – This chapter describes the development of a fluorescence-based process monitoring approach that was suitable for rapidly characterizing major NOM membrane foulants such as humic- and protein-like matter as well as colloidal/particulate matter present in natural water. The approach was also found suitable for identifying high membrane fouling events, which resulted from changes in the performance levels of different pre-treatment stages, in

drinking water treatment operations. This chapter has been published in the journal *Water Research* (Peiris *et al.*, 2010).

Chapter 9 – This chapter explains a novel technique involving an extension of the approach presented in Chapter 8, for understanding reversible and irreversible fouling behaviour in UF-based membrane treatment of drinking water using fluorescence measurements. This chapter has been accepted for publication in the *Journal of Membrane Science* (Accepted date: March 31, 2010; DOI: 10.1016/j.membsci.2010.03.047).

Chapter 10 – This chapter provides a logical continuation of the work presented in Chapter 9 in which development of a fluorescence-based modeling approach for fouling prediction and optimization of UF membrane processes for drinking water treatment is demonstrated. Parts of this chapter have been accepted to the *Invited Energy Session at the 9th International Symposium on Dynamics and Control of Process Systems (DYCOPS)* to be held in Belgium from July 7 – 9, 2010 (Submission date: November 12, 2009). Other parts of this chapter have been prepared for submission to the *American Institute of Chemical Engineers (AIChE) Journal*.

Chapter 11 – This chapter highlights the overall conclusions and recommendations resulting from the research presented in this thesis.

CHAPTER 2

Theoretical Background

2.1 Application of Membranes in Drinking Water Treatment

Membrane-based treatment processes are increasingly being used for the production of drinking water from surface and ground water sources. Incorporation of this technology allows effective removal of particulate matter, pathogenic organisms, salinity and NOM present in water (Thorsen, 1999) as well as requiring a smaller footprint than other approaches in the treatment facility. Membrane systems are operated in dead-end and cross flow modes and Figure 2-1 shows a schematic of these two modes of operation.

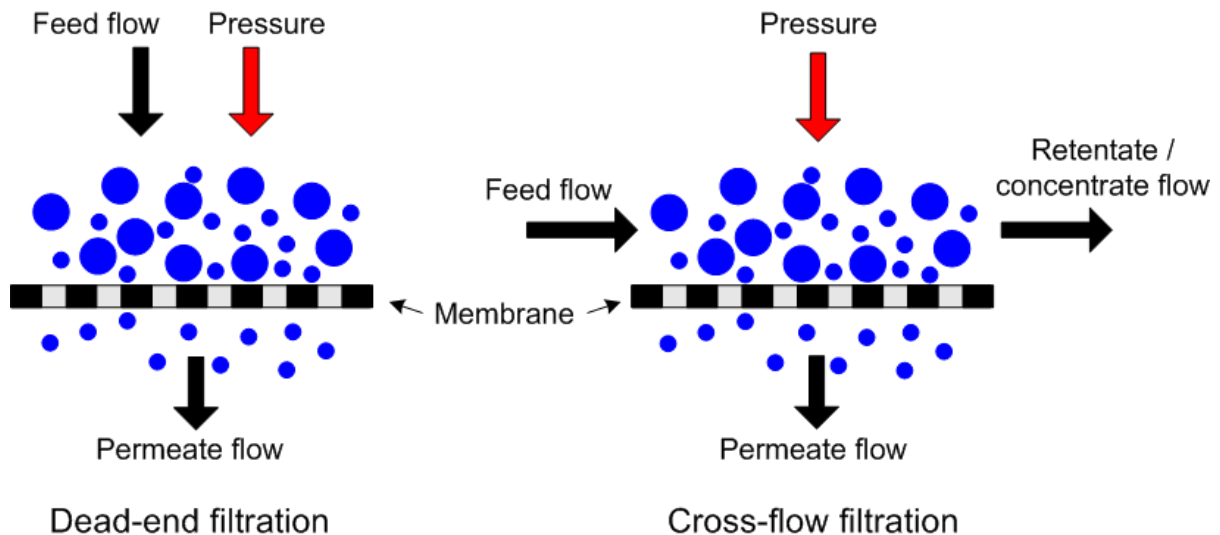


Figure 2-1 Schematic representation of the dead-end and cross flow filtration operations (adopted with modification from <http://www.hydrotech.cn/English/mofenli.asp>).

Different types of membrane filtration systems are used to achieve different treatment objectives. For example, microfiltration (MF) and UF membranes are typically used in drinking water treatment operations to replace granular media filtration (Freeman and William, 2005) and as barriers to particulate/colloidal matter present in water (Kimura *et al.*, 2005 and 2006; Fabris *et al.*, 2007). MF and UF of natural water can also reduce NOM and pathogenic organisms in water when used in combination with other pre- and post membrane treatment methods like coagulation and disinfection in water treatment systems (Fiksdal and Leiknes, 2006). In addition, UF membrane systems are also used as a pre-filtration step ahead of NF and reverse osmosis (RO) membrane systems to remove coarser particulate/colloidal material that can potentially contribute to fouling of NF and RO membranes (Glucina *et al.*, 2000). NF and RO technologies, on the other hand, are able to remove a wide range of contaminants such as DBP, synthetic organic compounds and pharmaceutical residues as well as salinity in the water (Glucina *et al.*, 2000; Hamed, 2005; Radjenović *et al.*, 2008).

2.2 Membrane Fouling in Drinking Water Applications

Implementation of membrane processes for drinking water treatment is constrained by the accumulation of materials (foulants) on the surface and/or in the pores of the membrane that decrease membrane permeability (Figure 2-2). This phenomenon is known as membrane fouling. Fouling increases operational costs by means of permeate flux decline and/or increased TMP requirements. In addition, frequent chemical cleaning of fouled membranes

leads to the deterioration of membrane performance, to shortened service life (Her *et al.*, 2007) and increased costs.

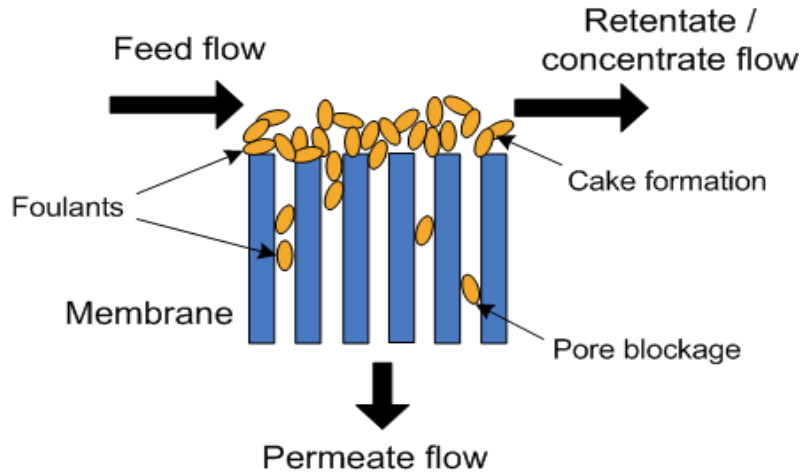


Figure 2-2 Schematic representation of membrane fouling.

2.2.1 Membrane Fouling Control Strategies

In practice, membrane fouling is controlled by implementing membrane cleaning operation schemes that include physical cleaning and chemical cleaning of fouled membranes. Membrane flux decline due to fouling that can be recovered by physical cleaning methods are usually identified as resulting from the reversible fouling of membranes. Membrane foulant components that are not easily removed by physical cleaning methods are thought to result from irreversible fouling; these fouling components are usually removed by chemical cleaning procedures (Kimura *et al.*, 2005). Physical cleaning methods include forward flushing, back-washing, vibrations and air flushing. In forward flushing, membrane surfaces are flushed with feed water or permeate generally at more rapid flow rates than during the production phase. During back-washing, permeate is forced in the opposite direction (i.e. from permeate side to feed side) under pressure. For air flushing, air is injected to the supply

pipe that carries the washing medium and employed during forward flushing or back-washing of membranes (Ebrahim, 1994). When physical cleaning methods do not improve the membrane flux sufficiently, chemical cleaning methods are applied. Typical chemical cleaning agents such as NaOH, KOH and NaOCl are used in this process to weaken the cohesion forces between the foulants and the adhesion forces between the foulants and the membrane surface (Trägårdh, 1989).

2.2.2 Major Membrane Foulants and Factors Affecting Fouling

NOM and colloidal/particulate matter present in water are the main factors responsible for membrane fouling in drinking water applications (Saravia *et al.*, 2006; Jermann *et al.*, 2007). A detailed description about NOM is presented in Section 2.5.

Membrane fouling is affected by many factors. These factors can be summarized into five main categories:

- (i) the properties of the NOM such as the bulk NOM concentration, humic/non-humic fractions, molecular weight distribution, size, hydrophobicity, charge density and isoelectric point (Jucker and Clark, 1994; Nilson and DiGiano, 1996; Hong and Elimelech, 1997; Cho *et al.*, 1999; Carroll *et al.*, 2000; Cho *et al.*, 2000; Lin *et al.*, 2000).
- (ii) the individual and combined effects of different NOM fractions, such as humic substances (HS), protein- and polysaccharide-like substances as well as colloidal/particulate matter present in natural water contribute to different

membrane fouling behaviour (Jermann *et al.*, 2008a and 2008b) involving reversible and irreversible membrane fouling.

- (iii) the solution properties such as pH, ionic strength, and metal ion concentration; metal ions such as Ca, Mg and Fe can interact with NOM to form fouling layers on the surface and/or in the pores of the membranes (Jucker and Clark, 1994; Braghetta *et al.*, 1997 and 1998; Hong and Elimelech, 1997; Combe *et al.*, 1999; Cho *et al.*, 2000).
- (iv) the properties of the membrane which include hydrophobicity, surface charge, surface roughness and porosity (Combe *et al.*, 1999; Kilduff *et al.*, 2000) and
- (v) the hydrodynamics of the membrane system (Hong and Elimelech, 1997; Braghetta *et al.*, 1998; Cho *et al.*, 1999; Taniguchi *et al.*, 2003).

Research activities on developing strategies for controlling membrane fouling should therefore be focused on the above factors. This research focused only on membrane fouling involving NOM and colloidal/particulate matter present in natural water.

2.2.3 Theoretical Basis of Membrane Fouling

Membrane fouling by NOM and colloidal/particulate matter can occur due to (i) concentration polarization that results from accumulation of retained foulants at the upstream surface of the membrane, (ii) pore blockage by solute adsorbed on the membrane surface or within pores and (iii) formation of a cake layer on the membrane surface, which represent resistance to flow in addition to the membrane resistance (Taniguchi *et al.*, 2003). Among these fouling mechanisms, concentration polarization effects have been shown to be

insignificant compared to pore blockage and cake formation for NOM related fouling in studies conducted using natural river water extracts (Cho *et al.*, 1999) and model foulants (Yuan and Zydney, 2000). The dominant fouling mechanism for colloidal/particulate-like material was shown to be cake formation (Jermann *et al.*, 2008b). In general, fouling components smaller than membrane pores can adsorb to membrane surfaces and reduce the cross-sectional area for flow while larger components can block pore entrances and contribute to cake formation (Taniguchi *et al.*, 2003). In addition, large fouling components (i.e. NOM aggregates and colloidal/particulate-like matter) could catalyze the fouling by smaller foulant species (Taniguchi *et al.*, 2003). In particular, interactions between larger NOM in the form of polysaccharides and colloidal/particulate-like matter can also contribute to simultaneous pore blocking and cake formation (Jermann *et al.*, 2008a).

Several studies have interpreted membrane fouling behaviour in terms of resistance-in-series models that account for membrane resistance resulting from each of these fouling mechanisms (Cho *et al.*, 1999 and 2000; Chang and Benjamin, 2003). A detailed description of these fouling mechanisms and their mathematical manifestations is presented in Tansel *et al.* (2000).

2.3 Membrane Fouling Modeling

2.3.1 Mechanistic Membrane Fouling Modeling Approaches

A number of membrane fouling modeling approaches have been developed to explain the fouling mechanisms of membrane-based separation systems. Mathematical modeling of

membrane fouling behaviour allows understanding how different fouling modes such as cake formation and pore blocking evolve during membrane filtration (Bowen *et al.*, 1995; Taniguchi *et al.*, 2003). In addition, membrane fouling modeling can also be used for assessment and prediction of different membrane fouling situations that have applications in membrane design, operation, maintenance, and process optimization (Lee *et al.*, 2009). Many studies have focused on modeling membrane fouling behaviour using mathematical/mechanistic modeling approaches (Bowen *et al.*, 1995; Tansel *et al.*, 2000; Chang and Benjamin, 2003; Taniguchi *et al.*, 2003; Bolton *et al.*, 2006).

These studies have focused mainly on dead-end and a few cross flow membrane filtration systems that did not involve membrane back-washing cycles that are typically applied in drinking water treatment systems. Also, these modeling approaches do not address the individual and combined contributions of different foulant fractions present in water on membrane fouling and therefore are not suitable for successfully predicting membrane fouling in drinking water applications. As a result, most often these models have limited use for practical applications and interpretation of field data. This can be attributed to the lack of knowledge of the various fouling mechanisms resulting from (1) the individual and combined contributions of different NOM fractions such as HS, protein- and polysaccharide-like substances, as well as colloidal/particulate matter present in natural water, (2) physical/chemical interactions of membrane foulants and (3) interaction between membrane surface and foulant components (Sahoo and Ray, 2006; Lee *et al.*, 2009).

2.3.2 Non-mechanistic Membrane Fouling Modeling Approaches

Due to the lack of understanding of different physicochemical phenomena involved in membrane fouling (discussed in Section 2.3.1) non-mechanistic approaches such as artificial neural networks (Delgrange-Vincent *et al.*, 2000; Cabassud *et al.*, 2002), empirical models (Shengji *et al.*, 2008) and genetic programming (Lee *et al.*, 2009) have been proposed for modeling of membrane fouling. In these studies, membrane fouling was related to long-term membrane feed water quality parameters and operational data including turbidity, temperature, dissolved organic carbon (DOC) content and TMP in pilot scale filtration studies.

However, these models are not able to capture the changes in the different membrane foulant fractions in water during the filtration nor can they relate different fouling behaviour with individual membrane foulant fractions. In addition, since for these empirical models the individual relations between the input variables and the modeled/predicted membrane fouling behaviour are not based on physical and engineering principles, the success of the resulting optimization strategies for fouling control is not always guaranteed.

2.4 Optimization of Membrane Filtration Processes for Drinking Water Treatment

Membrane filtration operations in drinking water treatment applications can be optimized to ensure maximum production of water per unit amount of energy consumed. This is achieved by implementing fouling control strategies (Seidel and Elimelech, 2002). One method of

controlling membrane fouling is the use of pre-treatment strategies such as coagulation, granular media filtration and chemical oxidation to improve the quality of the membrane feed water. In addition, for a given membrane feed water quality, membrane back-washing (sometimes assisted with air flushing) is widely considered as the most efficient technique for fouling minimization (Smith *et al.*, 2006). For optimization of membrane filtration processes in both drinking and wastewater treatment applications, the membrane back-washing time interval and the duration of the back-washes have been usually considered as control parameters (Konieczny, 2002; Smith *et al.*, 2005 and 2006; Zondervann *et al.*, 2007; Yigit *et al.*, 2009).

Most of these reported optimization approaches for membrane filtration were based on instantaneous membrane flux or TMP measurements made during membrane filtration. The optimization approaches that are based on model predictions into the future are limited. Thus, early predictions of the extent of reversible and irreversible membrane fouling for given raw water serving as feed to the membrane operations offer great potential for implementing efficient optimization strategies for membrane filtration. This research focused on developing such an optimization approach for minimizing membrane fouling as explained in Chapter 10.

2.5 Natural Organic Matter

Aquatic NOM contains a wide range of structurally complex compounds that result from prolonged degradation of plant and animal materials. The major fraction of NOM, which comprises over 50% of the DOC content, is composed of HS-like matter (Thurman, 1985).

Many researchers have suggested that HS are the most detrimental NOM-foulants, causing irreversible fouling by membrane adsorption and pore plugging (Jucker and Clark, 1994; Clark and Lucas, 1998; Combe *et al.*, 1999; Jones and O'Melia, 2000; Aoustin *et al.*, 2001). HS are operationally divided into humic acids (HA), fulvic acids (FA) and humus. Both HA and FA are present in water, with FA being the predominant part of HS (Huck, 1999). HS contain both aromatic and aliphatic components with mainly carboxylic, phenolic, alcoholic hydroxyl and keto functional groups (Thurman, 1985). Protein, amino acids, transphilic acids and polysaccharides fall into the non-humic part of NOM which account for 20 - 40% of the DOC content in natural waters (Fan *et al.*, 2001). Preferential fouling by these non-humic NOM components over HS occurs in some instances, depending on the hydrophilicity of the membranes (Kaiya *et al.*, 2000; Her *et al.*, 2007). These major NOM fractions occur largely in dissolved form (Thurman, 1985) while a minor portion of protein- and polysaccharide-like matter is found in colloidal/particulate-like matter form (Amy, 2008). Colloidal/particulate matter in natural water is comprised of NOM-like material and inorganic components (Aoustin *et al.*, 2001; Jermann *et al.*, 2008b).

2.5.1 Natural Organic Matter as a Precursor for Disinfected By-product Formation

NOM contains precursors that can give rise to the formation of DBP following the disinfection/oxidation of drinking water (Marhaba and Washintgton, 1998; Marhaba and Kochar, 2000). The potential of DBP formation for different NOM fractions varies considerably. It is suspected that most precursors to DBP formation are HS, which are the main constituents of NOM. DBP are known for their carcinogenic and mutagenic potential and are mainly found in the form of trihalomethanes, haloacetic acids, haloacetonitrile,

cyanogen and halopictins (Marhaba and Kochar, 2000). Due to these concerns, formation of DBP by NOM has been under close scrutiny since the discovery of their presence in drinking water.

2.6 Significance of NOM Characterization

Characterization of NOM, including colloidal/particulate matter, in natural water is essential for the understanding and identification of their changes that occur during different stages of drinking water treatment. The information obtained from the characterization of these components in the membrane feed and membrane permeate proved to be invaluable for the development of fouling control and optimization strategies as demonstrated in Chapter 10. In addition, such characterizations performed at different pre-membrane treatment stages such as granular media filtration, biofiltration and coagulation can also be helpful for implementing monitoring and control strategies that focus on ensuring membrane feed water with low foulant content as demonstrated in Chapter 8. NOM that may pass through membranes also contributes to DBP formation and increases the need for maintaining an effective disinfectant (chlorine) residual concentration during treatment and distribution (Ring *et al.*, 2004). Therefore, NOM characterization at post membrane stages could be helpful in assessing the DBP formation potential in water and estimating the extent of disinfection that is needed to maintain an effective residual disinfectant content in drinking water distribution systems.

2.7 Characterization of Natural Organic Matter Foulants

The composition of NOM is very complex in nature and therefore presents some unique challenges when using more established techniques for the characterization of organic matter (Christy and Egeberg, 2000). Reported methods for the characterization of NOM have mostly focused on traditional analyses such as light absorptivity, DOC concentration and aromaticity due to the difficulty of detailed structural analysis (Her *et al.*, 2003). Direct application of these techniques, on un-fractionated water, are however not capable of adequately examining the composition of the NOM. For example, the UV absorbance spectrum of NOM is virtually featureless, except for a general trend of decreasing absorbance with increasing wavelength due to the presence of a broad diversity of UV-absorbing chromophores with overlapping absorbance bands (Li *et al.*, 2006). Nevertheless, characterization of NOM based on correlations of UV absorbance with the aromatic content of NOM has been reported (Christman *et al.*, 1989; Traina *et al.*, 1990; Yu-Ping *et al.*, 1994; Peuravuori and Pihlaja, 1997; Croue *et al.*, 2000; Zhou *et al.*, 2001). The specific absorbance, calculated in terms of a ratio of UV absorbance to DOC content at 254 nm or 280 nm, has been correlated with aromaticity, the number and the weight averaged molecular weights (MW_n and MW_w), and has been proposed as a measure of NOM in water. Traditionally, the specific absorbance at 254 nm is termed as $SUVA_{254}$ and the same ratio at 280 is referred to as the molar extinction coefficient (ϵ_{280}) of NOM (Li *et al.*, 2006).

Studies on the structural characteristics of NOM, on the other hand, demand more advance analytical approaches like nuclear magnetic resonance (^{13}C NMR) spectroscopy, infrared (IR) spectroscopy, differential thermal analysis, modulated differential scanning calorimetry,

and pyrolysis-gas chromatography/mass spectrometry (Wilson *et al.*, 1999; Smeulders *et al.*, 2000). In addition, Fourier transform infrared spectroscopy (FTIR) is known to provide information related to the presence of specific functional groups of various isolates of NOM (Lee *et al.*, 2006; Gray *et al.*, 2007). Liquid Chromatography - Organic Carbon Detection (LC-OCD), a relatively new and more sensitive analytical technique, is also available for obtaining both quantitative and qualitative information on NOM (Huber and Frimmel, 1992).

Unfortunately, most of these characterization methods demand concentration and/or fractionation of NOM before analysis due to their low sensitivity to the low concentration levels found in raw water (Croue, 2004). Aside from their unsuitability for on-line applications, all of these above mentioned methods have the disadvantages of being time consuming, expensive, complex and are based on difficult analyses that require operator expertise (Her *et al.*, 2003). In this context, fluorescence spectroscopy is becoming an increasingly popular method for characterizing NOM and shows good potential for on-line monitoring, as minimal sample pre-treatment and preparation is required, high instrumental sensitivity is available (Coble *et al.*, 1990; Coble, 1996) and the technique is non-destructive in nature (Mobed *et al.*, 1996).

2.8 Principles of Fluorescence Spectroscopy

Fluorescence is a form of luminescence, which is the emission of light from any substance. The type of luminescence that occurs when molecules are excited by ultraviolet or visible light photons is termed as photoluminescence, and is formally divided into two categories:

fluorescence and phosphorescence, depending on the electronic configuration of the excited states and the emission pathways (Herman *et al.*, 2003; Lakowicz, 2006).

2.8.1 Fluorophores

The substances which contain molecules capable of undergoing electronic transitions that result in fluorescence are known as fluorescent probes or fluorochromes. Fluorochromes that are conjugated to a larger macromolecule (such as a nucleic acid, lipid, enzyme, or protein) through adsorption or covalent bonds are termed fluorophores (Herman *et al.*, 2003). Fluorophores generally contain electrons which are present in conjugated systems (Lakowicz, 2006). A chemically conjugated system is a system of atoms covalently bonded with alternating single and multiple (e.g. double) bonds (e.g., $C=C-C=C-C$) in a molecule of an organic compound. The presence of nitrogen in aromatic ring formations, carbonyl groups ($C=O$), amine groups ($C=N$), vinyl groups ($C=C$), or anions will also suffice for maintaining conjugation (Lakowicz, 2006).

2.8.2 Theory of Fluorescence

When a substance having a fluorophore (or a combination of fluorophores) is irradiated by a light source with a wide spectrum of wavelengths, the fluorophore(s) in the substance will absorb light and reach various excited electronic and vibrational states. Within each of these electronic states there are various vibrational states. Depending on the electronic configuration that results due to the conjugation in the fluorophores, some of these transitions will have a much higher degree of probability than others, and when combined, will constitute the unique absorption spectrum of the substance. The wide range of photon

energies associated with absorption transitions in fluorophores causes the resulting fluorescence spectra to appear as unique broad bands of spectrum (Herman *et al.*, 2003). In this way, different electronic configurations that are unique to different fluorophores will generate fluorescence spectra with unique features. Processes which occur between the absorption and emission of light are usually illustrated by a Jablonski diagram. A typical Jablonski diagram is shown in Figure 2-3. The ground, first and second electronic states are depicted by S_0 , S_1 and S_2 , respectively. At each of these electronic energy levels, the fluorophores can exist in a number of vibrational energy levels denoted by 0, 1, 2, etc. Transitions between states are depicted as vertical lines to illustrate the instantaneous nature of light absorption. A detailed description of the theory of fluorescence spectroscopy is found in Lakowicz (2006).

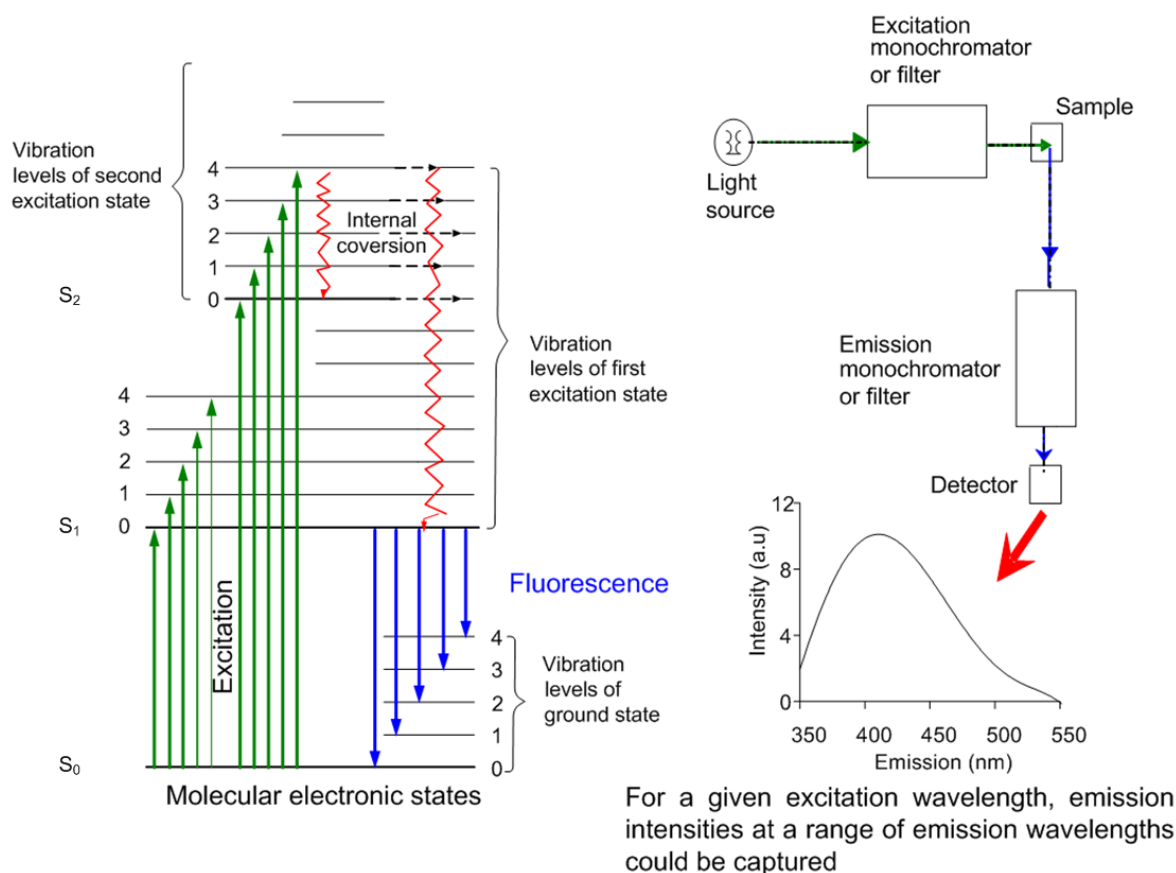


Figure 2-3 A simplified Jablonski diagram that illustrates the processes that occur between absorption of light and generation of fluorescence signals (adopted with modifications from Herman *et al.*, 2003).

NOM fractions present in water contain fluorophores with different types of functional groups resulting in unique fluorescent properties (Sharpless and McGown, 1999; Chen, J. *et al.*, 2003). Therefore, by examining the differences in the fluorescence spectral properties due to the presence of these functional groups, characterization of different NOM fractions is possible.

2.9 Characterization of NOM by Fluorescence Spectroscopy

Fluorescence spectroscopic techniques have been increasingly utilized for studying the characteristics of organic matter in natural water. To produce meaningful fluorescence intensity readings the following fluorescence analysis approaches are available:

- (i) single emission scan at fixed excitation wavelengths,
- (ii) synchronous scan at a constant offset wavelength between excitation and emission wavelengths, and
- (iii) excitation-emission matrix (EEM) analysis (Chen *et al.*, 2003).

Although single emission scan and synchronous scan approaches have been used for NOM characterization, the fluorescence EEM analysis method has been widely accepted to provide more detailed spectral information of all the features existing within a selected spectral range for NOM in natural water (Mopper and Schultz, 1993; Coble, 1996; Matthews *et al.*, 1996; Del Castillo *et al.*, 1999; Mounier *et al.*, 1999; Parlanti *et al.*, 2000 and 2002; Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003; Sierra *et al.*, 2005). This method captures a large number of fluorescence intensity readings at different combinations of excitation and emission wavelengths forming a matrix of fluorescence intensities. This approach has also been observed to provide better sensitivity compared to the single emission scan approach (Peiris *et al.*, 2010).

2.9.1 Characterization of Particulate/colloidal Matter using Rayleigh Scattering

Light scattering is known to provide information related to the colloidal/particulate matter present in water (Wyatt, 1993; Stramski and Wozniak, 2005). First order Rayleigh scattering

(FORS) regions, which occur when excitation wavelength (E_x) is approximately equal to emission wavelength (E_m), and second order Rayleigh scattering (SORS) regions, which occur when $E_m \approx 2.E_x$, can be captured within fluorescence EEMs (Rinnan and Andersen, 2005). These Rayleigh scattering regions (RS) can therefore be used to obtain information related to particulate/colloidal matter present in water as shown in Chapter 8.

2.9.2 Effect of the Solution Properties, the Season and the Source of Water on Fluorescence Spectral Characteristics

Due to the presence of multiple pH-dependent fluorophores in NOM, characterization of NOM by fluorescence spectroscopy is affected by varying pH conditions (Pullin and Cabaniss, 1995). The results reported by Pullin and Cabaniss (1995) indicated that fluorescence spectra of HS- and protein-like matter were affected by pH. The effect of ionic strength was however often observed not to be significant on fluorescence spectral characteristics of HS- (Mobed *et al.*, 1996; Fu *et al.*, 2004) and protein-like matter (Kwaambwa and Maikokera, 2007).

Fluorescence EEM spectra of natural waters are also known to be affected by the source and seasonal effects. The following examples from the literature illustrate these effects: (i) changes in the relative fluorescence intensity of peaks due to seasonal variation of NOM from rivers have been reported (Alberts *et al.*, 2001); (ii) Yan *et al.* (2000) used the fluorescence EEM approach as a tool to identify the source of water intrusion from multiple sources; and (iii) possibilities of discriminating soil-derived and aquatic-derived HS and HA

and FA derived from the same source using fluorescence spectroscopy was observed (Mobed *et al.*, 1996).

2.10 Techniques that Characterize the Properties of Fluorescence Spectra

Most of the reported fluorescence EEM-based techniques that characterize the spectral properties of NOM have considered only major fluorescence peak positions (peak-picking method) out of the entire fluorescence spectrum that contain thousands of wavelength-dependent fluorescence intensity data points (Chen *et al.*, 2003). These techniques lack the ability to capture the heterogeneity of the different NOM fractions in water and therefore do not provide accurate characterization of specific NOM fractions. The importance of analyzing all the fluorescence intensities captured at different excitation- and emission-wavelength combinations has been emphasized in many studies. This way of analyzing the fluorescence EEMs can provide more accurate characterization of NOM foulants present in water as opposed to the more commonly used method in which only a few excitation-emission coordinate pairs (i.e. main peaks) are examined (Persson and Wedborg, 2001; Chen *et al.*, 2003; Stedmon *et al.*, 2003; Boehme *et al.*, 2004).

Chen *et al.* (2003) have introduced a new approach termed fluorescence regional integration (FRI) that allowed the analysis of the full fluorescence spectrum. In this approach, EEMs of dissolved organic matter (DOM), which is the dissolved part of NOM, was operationally divided into five different regions based on established peak locations of different DOM fractions. The quantification of different DOM fractions by using FRI was then based on the

postulate that integration beneath EEMs within selected regions would represent the cumulative fluorescence response of DOM with similar properties (Chen *et al.*, 2003). It could be argued that this seemingly elegant method may need division of EEMs into smaller regions to capture all the unique spectral characteristics of different NOM fractions for increased accuracy. Also, there are no systematic ways to determine the sub-regions to be used for these calculations.

In this context, multivariate statistical tools have been considered as good candidates for capturing the spectral information in the full fluorescence spectrum. Parallel factor (PARAFAC) analysis (Stedmon *et al.*, 2003) is such an approach that has been used for fluorescence EEM-based characterization of NOM in water. The limitation with the PARAFAC approach is that it is not suitable for the quantitative analyses of different NOM fractions. Principal component analysis (PCA), another multivariate statistical data analysis method, has also been used to perform similar characterizations. In addition, partial least squares (PLS) regression, a multivariate statistical technique, has been successfully used for quantification of protein mixtures by using their fluorescence EEMs in a separate study (Elshereef *et al.*, 2006).

2.11 Multivariate Data Analysis of Fluorescence Excitation-emission Matrices

In this research, PLS was used as a tool to analyze all the intensity data points in the full fluorescence EEM for developing calibration models which correlate the intensities captured

within the full fluorescence EEMs of water samples containing different concentration levels of model and natural NOM fractions with their corresponding DOC concentrations. In addition, PCA was used to extract principal components (PCs) that contained information relevant to major membrane foulant fractions such as HS-like, protein-like and colloidal/particulate matter from fluorescence EEM measurements collected at different stages in a drinking water treatment process. This approach was useful in developing fluorescence-based membrane fouling prediction and optimization approaches as will be discussed in Chapter 10.

2.11.1 Partial Least Squares (PLS) Regression

PLS methodology has been applied successfully in application areas such as (i) quantitative structure-activity relationship modeling, (ii) multivariate calibration, and (iii) process monitoring and optimization (Eriksson *et al.*, 2001). PLS regression analysis is a multivariate statistical technique that allows comparison between multiple response variables and multiple explanatory variables. In the case of this research, the multiple response variables are the DOC contents of different NOM fractions present in water and the different excitation-emission wavelength combinations in the fluorescence EEMs of these water samples are identified as multiple explanatory variables. PLS derives its usefulness from its ability to analyze data with noisy, collinear and even incomplete variables in response (Y) and explanatory/predictor (X) spaces (Eriksson *et al.*, 2001). PLS analysis can perform regression and classification tasks as well as dimension reduction techniques in order to find stable and reliable solutions (Field, 2000). Therefore, it has increasingly gained popularity in applications related to science, especially chemistry and chemometrics, where there is a

problem with a high number of correlated variables and a limited number of observations (Eriksson *et al.*, 2001). Correlated data is a common occurrence when dealing with spectral data where the measurements acquired at different excitation-emission wavelengths are highly correlated with each other.

The underlying assumption of the PLS method is that the observed or measured data could be mathematically described by a small number of latent (not directly observed or measured) variables (Rosipal *et al.*, 2006). In general, PLS tries to extract these latent variables (LVs) – also known as latent components – accounting for as much of the manifest covariance as possible by projecting the observed data onto these LVs. For this reason, the acronym PLS also stands for projections to latent structures by means of partial least squares.

2.11.1.1 Significance of PLS Compared to Other Classical Methods of Statistics

Common classical methods of statistics, such as multiple linear regression (MLR), canonical correlation, linear discriminate analysis, analysis of variance (ANOVA) and maximum likelihood methods function on the underlying assumptions that the input variables (X) are independent, exact and the residuals between the predicted and measured response variables (Y) are randomly distributed. With PLS however, these underlying assumptions are not needed. That is, in PLS, X-variables need not be independent or exact and residuals may not be randomly distributed. In addition, PLS is also capable of handling noisy, collinear and even incomplete data sets providing much leverage for complex data analysis (Eriksson *et al.*, 2001).

2.11.1.2 Comparison Between PLS and Other Multivariate Regression Methods

In addition to PLS, canonical correspondence analysis (CCA), redundancy analysis (RDA) and principal component regression (PCR) are three common multivariate regression methods; however PLS, was identified as the most suitable approach for this research due to the following reasons:

- PLS is a biased regression method whereas CCA and RDA are based on unbiased regression (Ter Braak *et al.*, 1995). Unbiased regression methods predict poorly if the predictor variables are very highly correlated (i.e. multicollinearity) as is the case with fluorescence EEM data. This is because multicollinearity “inflates” the variance of the parameter estimates in an unbiased regression. Biased regression methods, on the other hand, estimate biased regression parameters so that variance of the parameter estimates are not as affected as in unbiased regression methods (Gunst *et al.*, 1981).
- Unlike PLS, if variables are highly correlated, the weights of the ordinations (LVs) in CCA and RDA become unstable and not interpretable (Ter Braak *et al.*, 1995). In contrast, PLS weights provide a logical basis to interpret the actual phenomenon captured in the regression (Eriksson *et al.*, 2001).
- The response data for CCA must be nonnegative; the data are abundances (e.g. counts or presence-absence) or compositional data, in the sense that only relative values are meaningful. In contrast, typical response data in PLS are quantitative, either positive

or negative, without special meaning attached to the value zero (Ter Braak *et al.*, 1995).

- Similar to PLS, PCR is able to handle multicollinearity in the predictor data and often can generate calibration models that give prediction errors as low as those of PLS. Unlike PLS, PCR generates LVs from the predictor data without using the information related to the correlations that exist between predictor and response data. Due to this reason, PCR-based calibration models use a larger number of LVs that are often difficult to interpret compared to the PLS-based calibration models (Esbensen, 2002).

2.11.1.3 PLS Regression Methodology

2.11.1.3.1 Pre-processing of Data

Prior to PLS analysis, raw data are processed to ensure that the data are symmetrically distributed and have a reasonably constant error variance. Three common data pre-processing methods used in multivariate data analysis are:

- (i) Mean centering – calculates the mean of a variable and subtracts the mean from each value of that variable
- (ii) Median centering – similar to mean centering except the reference point is the median of each variable rather than the mean
- (iii) Auto-scaling – variables are mean centered and scaled to unit variance

In PLS regression, variables are usually auto-scaled before the analysis so that all variables are given an equal importance. However, if prior knowledge about the relative importance of variables in X and Y is available, variables could be scaled accordingly by giving important variables a slightly higher weight than that corresponding to unit variance scaling; similarly, unimportant variables should be given a slightly lower weight (Eriksson *et al.*, 2001).

2.11.1.3.2 Theory

The PLS regression consists of simultaneous projections of both X and Y spaces onto low dimensional hyper-planes – also known as LVs – so that the analysis can meet the following objectives:

- To approximate the X and Y spaces and
- To maximize the correlation between X and Y spaces.

A general PLS regression model that accomplishes these objectives can be expressed as:

$$X = TQ' + E = \sum_{i=1}^m t_i q_i + E_o \quad (2.1)$$

$$Y = UC' + F = \sum_{i=1}^m u_i c_i + F_o \quad (2.2)$$

The inner relation that explains the correlation between X and U (or in effect X and Y) is given by:

$$U = TW \quad (2.3)$$

Where m is the number of LVs used in capturing the correlation between X and Y data sets. The scores corresponding to the i^{th} LV of X and Y space are identified as vectors \mathbf{t}_i and \mathbf{u}_i , respectively and summarized in the i^{th} column of T and U matrices, respectively. The scores in vector \mathbf{t}_i can be interpreted as projections of the predictor data (e.g. fluorescence spectral variables) to the i^{th} LV in X space. Similarly, vector \mathbf{u}_i can be interpreted as projections of the response data (e.g. DOC content of NOM fractions) to the i^{th} LV in Y space. The derivation of vectors \mathbf{t}_i and \mathbf{u}_i , in terms of projecting the raw data to LVs are illustrated in Figure 2-4. Vectors \mathbf{q}_i ($i = 1, 2, \dots, m$) are known as loading values that express the relationship between T and X-space whereas \mathbf{c}_i ($i = 1, 2, \dots, m$) vectors relate the U and Y space. The loading values corresponding to the i^{th} LV in X and Y spaces are summarized in the i^{th} column of Q and C matrices, respectively

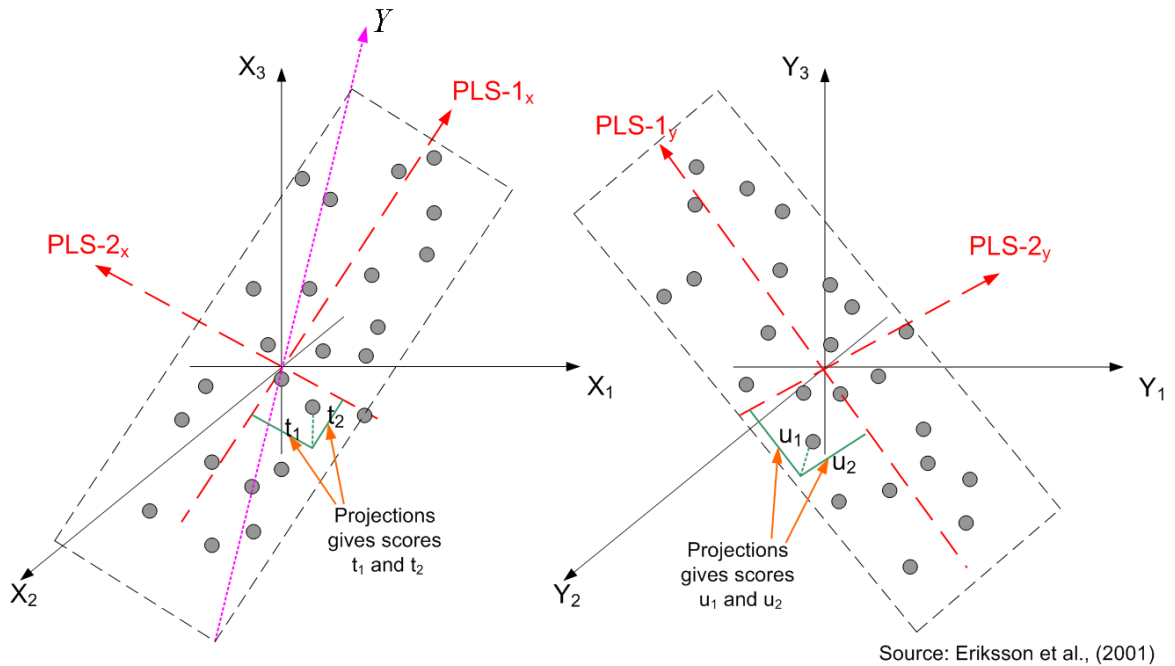


Figure 2-4 Schematic representation of the derivation of PLS components and the corresponding scores in multivariate explanatory and multivariate response spaces. PLS-1 and PLS-2 represent 1st and 2nd PLS components. Only three variables, each for X and Y spaces, are shown for clarity.

In addition, the PLS algorithm also derives a W weight matrix that explains the correlation between X and U (i.e. in effect X and Y -space) as illustrated in Equation 2.3. The W weight matrix represents how the X -variables are linearly combined to form any score vector t_i when capturing the correlation between X and Y . Hence, by examining the W matrix, one could understand which original variable in X space would dominate the new LVs (Eriksson *et al.*, 2001). The variation in the data that was left unexplained by the PLS modeling is given by the E_o and F_o residual matrices. Figure 2-5 illustrates the matrix relationship in PLS regression. More detailed descriptions of the PLS methodology is found in Eriksson *et al.* (2001) and Esbensen, (2002).

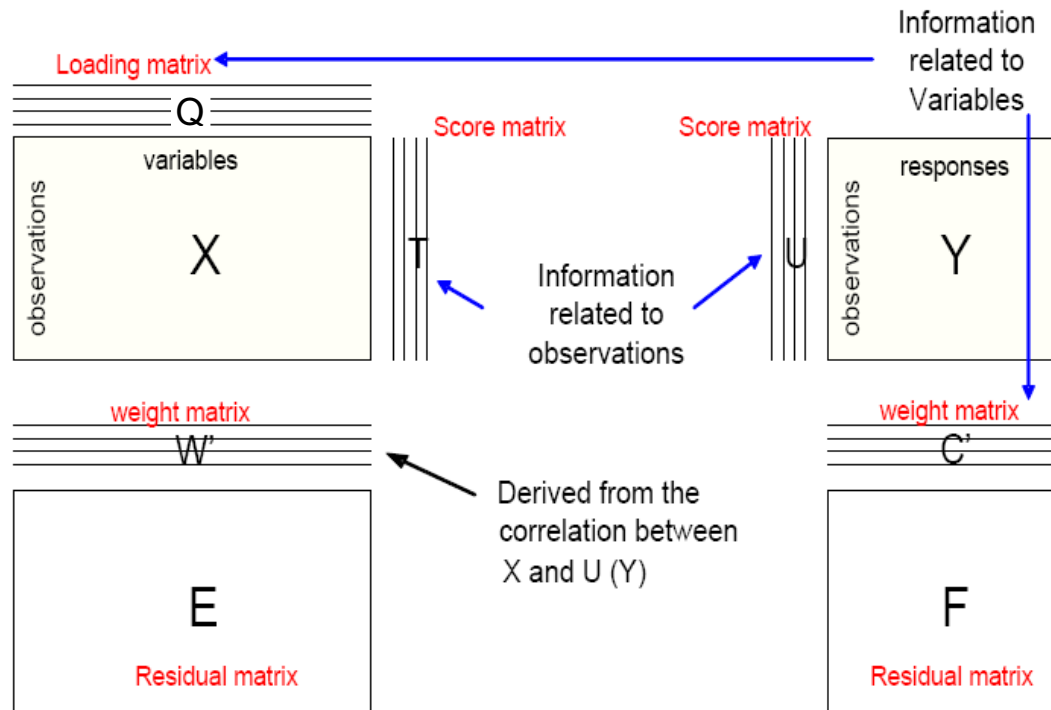


Figure 2-5 The matrix relationship in PLS regression. The score, weight and loading matrices are derived during the development of the PLS regression model. Source: Eriksson *et al.* (2001).

When the PLS components or LVs are selected for developing PLS regression models, one should avoid *over-fitting* of the data. As the number of PLS components used in a PLS model increases, the *goodness of fit* (R^2), which corresponds to the calibration data set, will also improve. However, including more than an optimum number of PLS components can also reduce the model's prediction accuracy when the regression model is used to predict independent data sets that were not used for model calibration. Therefore PLS algorithms are often included with cross validation techniques that can be used to find the optimum number of PLS components (or LVs) that are needed in modeling a given scenario. Cross validation ensures the optimal balance between the *goodness of fit* and the predictive ability of PLS models (Eriksson *et al.*, 2001).

2.11.1.3.3 Finding the Optimum Number of PLS Components (or LVs)

Cross-validation is performed with the aim to determine the optimal model complexity and the minimum number of LVs that describes the underlying relationship between the X-space and Y-space. It is a procedure where the available data set is split between training (calibration data set) and test sets. The prediction residual error on the test samples is determined as a function of the number of PLS components/LVs retained in the regression model formed with the calibration data. The procedure is usually repeated several times, with each sample in the original data set being part of the test set at least once. The total prediction error over all the test sets as a function of the number of PLS components is then used to determine the optimum number of PLS components, i.e. the number of PLS components which produces minimum prediction error. There are several cross validation methods that can be used and each of them performs differently depending on the data set being modeled.

Most commonly used methods are: (i) leave-one-out method, (ii) random subsets, (iii) contiguous block and (iv) venetian blinds method (Wise *et al.*, 2004).

2.11.2 Principal Component Analysis

PCA, like PLS, is a multivariate statistical tool that is suitable for analyzing a large number of highly collinear variables such as spectral variables in fluorescence EEMs. Unlike PLS, PCA only analyzes the predictor data matrix (i.e. only fluorescence EEM data) and is used for identifying patterns in data, and to highlight their similarities and differences (Eriksson *et al.*, 2001; Esbensen, 2002).

2.11.2.1 PCA Analysis Methodology

2.11.2.1.1 Pre-processing of Data

Raw data needs to be pre-processed prior to performing PCA. This can be achieved by following the same data pre-processing methods described in Section 2.11.1.3.1.

2.11.2.1.2 Theory

PCA extracts a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and mathematically represented by linear combinations of original variables in the data set (or data matrix X). These new variables, referred to as PCs, are calculated to capture much of the variance present in the X matrix (Wold *et al.*, 1987; Eriksson *et al.*, 2001) and therefore can describe major trends in the original data set. PCA decomposes the data matrix X as the sum of the outer product of vectors \mathbf{s}_i and \mathbf{p}_i plus a residual matrix E which contains the variation in the data that was left unexplained as presented in Equation 2.4.

$$X = \sum_{i=1}^n s_i \cdot p_i + E \quad (2.4)$$

Where n is the number of samples in the X data set. The s_i vectors are known as scores (i.e. values) on the PCs (i.e. new variables) extracted by PCA. The p_i vectors are known as loadings and contain information on how the variables relate to each other (Wold *et al.*, 1987; Eriksson *et al.*, 2001). The scores (s_i) generated by PCA can be interpreted as projections of original variables to a new space spanned by the PCs (i.e. when the original variables are transformed to PCs, each variable in the X matrix is projected on to the PC space). The co-ordinates in this PC space are the scores. The set of scores corresponding to a particular PC can be plotted against another set of scores corresponding to another PC and such plots are called score plots. Generally, the score plots are built on the basis of the first two PCs since these usually explain most of the variability in the data. The PCs in the PC space are related to the variables in the X matrix (i.e. original variable space) by loadings (Persson and Wedborg, 2001). By examining the loading values related to each PC, it is possible to understand which original variables in the X matrix are better explained by each PC.

2.11.2.1.3 Selecting Statistically Significant PCs

When PCA is performed one must determine the number of PCs to retain in the model. This is achieved by one or both of the following two approaches (Wise *et al.*, 2004).

- (i) Examining the eigenvalues related to each PC in the model. Occurrence of consecutively smaller eigenvalues followed by comparatively larger eigenvalues usually indicates the presence of statistically insignificant PCs.

- (ii) Performing cross validation to identify the minimum number of statistically significant PCs that can be used to reconstruct the original data set with minimum reconstruction error. Different cross validation methods described in Section 2.11.1.3.3 can be used.

CHAPTER 3

Acquiring Reproducible Fluorescence Spectra of Dissolved Organic Matter at Very Low Concentrations^{*}

Overview

A method that would allow for fast and reliable measurements of dissolved organic matter (DOM), both at low and high concentration levels would be a valuable tool for on-line monitoring of DOM. This could have applications in a variety of areas including membrane treatment systems for drinking water applications which is of interest to our group. In this study, the feasibility of using fluorescence spectroscopy for monitoring DOM at very low concentration levels was demonstrated with an emphasis on optimizing the instrument parameters necessary to obtain reproducible fluorescence signals. Signals were acquired using a cuvette or a fibre optic probe assembly, the latter which may have applications for on-line or in-line monitoring. The instrument parameters such as photomultiplier tube (PMT) voltage, scanning rate and slit width were studied in detail to find the optimum parameter settings required. The results showed that larger excitation and emission slit widths were preferred, over larger PMT voltage or lower scanning rates, to obtain reproducible and rapid measurements when measuring very low concentration levels of DOM. However, this

* Peiris, B. R. H., Budman, H., Moresoli, C., Legge, R. L. 2009. Acquiring reproducible fluorescence spectra of dissolved organic matter at very low concentrations. *Wat. Sci. And Technol.* **60** (6), 1385-1392.

approach should be implemented with caution to avoid any reduction of the signal resolution.

Keywords: Dissolved organic matter (DOM); drinking water treatment; fluorescence spectroscopy; humic substances; measurement noise

3.1 Introduction

Interest in dissolved organic matter (DOM) is increasing in the area of water treatment, particularly as it relates to membrane fouling, monitoring and control of disinfected by-products and issues arising due to interactions with other environmental contaminants (Mobed *et al.*, 1996; Marhaba and Kochar, 2000). Most methods available for characterizing DOM rely on traditional analysis techniques such as light absorptivity, dissolved organic carbon (DOC), aromaticity and fractionation using XADTM absorbent resins although characterization is difficult due to the complex structural aspects of DOM (Her *et al.*, 2003). Some studies have been performed on molecular weight (MW) fractions of DOM using nuclear magnetic resonance (¹³C NMR), pyrolysis-GC/MS, IR and thermogravimetric methods (Wilson *et al.*, 1999). Aside from the fact that these methods are time consuming, expensive, complex and require operator expertise (Her *et al.*, 2003) they are also not amenable to on-line or in-line applications for monitoring DOM.

In this context, fluorescence spectroscopy is becoming an increasingly popular method for characterizing DOM due to the requirement for minimal sample pre-treatment and preparation, ability to interpret DOM fluorescence properties, high instrumental sensitivity and the non-destructive nature of the technique (Peiris *et al.*, 2008). Chen *et al.* (2003)

identified various fluorescence spectroscopic techniques, out of which, the emission-excitation matrix (EEM) analysis method was identified as providing an overall view of all features of DOM existing within a selected spectral range (Sierra *et al.*, 2005). This mode of fluorescence data acquisition has been used satisfactorily in water analysis (Her *et al.*, 2003; Sierra *et al.*, 2005) and, with these advantages, is promising as an on-line or in-line monitoring approach (Peiris *et al.*, 2008).

The DOM content in drinking water is generally modest with the most significant component being humic substances (HS) and some minor levels of polysaccharides and protein related substances. HS may represent up to 90% of the DOM with the concentration varying from 0.5 to 30 mg-C/L in streams and rivers (Thurman, 1985). Especially where membrane treatment methods are being considered, the DOM content at the permeate side can be very low resulting in challenges associated with the characterization of the DOM content in the permeate. In such cases, most of the analytical techniques described above require that the DOM content be concentrated by means of extraction or other physical pre-treatment methods prior to analysis. In this paper, we have performed a comprehensive study on the instrument parameters that are necessary in order to obtain reproducible fluorescence signals for DOM at very low concentration levels. The main emphasis was to obtain reproducible spectra at low concentration levels for application in areas such as characterizing the DOM matter at the permeate side for a membrane-based water treatment process. Even though fluorescence spectrofluorometer literature, including instrument manuals, provides generally useful information on how the instrument parameters may be changed to obtain reproducible fluorescence readings, manipulation of these instrument parameters, especially at very low

concentration levels, has been rarely demonstrated. This may explain in part the reason for limited direct applications of fluorescence spectroscopy in the characterization of very low DOM concentration levels such as the levels found in nanofiltration (NF) permeates. In this context by using the optimum instrument parameters settings determined in this study, we have managed to characterize DOM in the NF permeate of drinking water (Peiris *et al.*, 2008). These findings also have value in developing a fluorescence-based sensor for an on-line approach for monitoring DOM content in drinking water treatment applications. For these reasons, this paper will serve as a useful reference for those interested in developing fluorescence spectroscopy as a tool to monitor low concentration DOM, for example in post water treatment strategies.

3.2 Methods

3.2.1 Water Samples and Preparation

Water samples were obtained on March 6th, 2007 from the Grand River in Kitchener, Ontario, Canada. Grand River water (GRW) samples were filtered using a 0.45 μm Nylon Acrodisc type filter to remove any particulate material present. The samples were then diluted to different dilution levels using Milli-Q (Millipore) water as the dilution medium. Different dilution levels spanned from no dilution to dilutions up to 10% of the original strength of the GRW. The value “10% dilution level” was selected to match the dissolved organic carbon (DOC) content of the permeate, when GRW was filtered through a nanofiltration (NF) system currently in use. During our preliminary studies, permeate with ~ 0.7 mg-DOC/L was obtained from GRW with an average 6.7 mg-DOC/L. The pH of the

GRW was ~ 8.2 so the pH of all diluted samples was adjusted to ~ 8.2 using high purity 0.1 M HCl acid (99.999%) or 0.1 M NaOH (99.998 %).

3.2.2 Fluorescence Analysis

Fluorescence excitation emission matrices (EEMs) of each sample were acquired using a Varian Cary Eclipse Fluorescence Spectrofluorometer (Palo Alto, CA) at ~ 25 °C. A Peltier multicell holder, which can accommodate cuvettes, and a Fluorescence Remote Read Fibre Optic Probe coupled to an Eclipse Fibre Optic Coupler with a 20 mm fluorescence probe tip, were used for signal acquisition. UV-grade polymethylmethacrylate cuvettes with four optical windows were used for cuvette-based analysis. The excitation and emission ranges used were 280-380 nm and 350-600 nm, respectively which fall in most reported ranges for HS (Chen *et al.*, 2003; Her *et al.*, 2003; Sierra *et al.*, 2005). Fluorescence emission spectra were obtained at different photomultiplier tube (PMT) voltages (V), scanning rates and emission and excitation slit widths. To eliminate Raman scattering due to water and to reduce other background noise, all spectra were subtracted from the spectra for Milli-Q water obtained under the same conditions. Data processing was performed using MATLAB 7.3.0 software (The Mathworks Inc., Natick, MA). Analysis of the reproducibility of the spectra was based on the measurement noise (or % error with respect to signal intensity) which was calculated using error at a 95% confidence level of the peak intensity measurement for all dilution levels.

3.3 Results and Discussion

DOM in surface and ground water consists of a variety of organic compounds with various fluorophores that are responsible for producing unique spectral fingerprints (Sierra *et al.*, 2005). HS and protein-related substances are the primary sources of fluorescence in DOM with HS being the dominant fluorescent component. However, with the low concentration levels of HS found in drinking water treatment applications, the weak fluorescence signals are compromised due to large measurement noise. This issue becomes more significant for on-line measurements acquired using fibre optic probes as signal strength in this approach is further attenuated in the fibre optic system. As a result, off-line measurements are often obtained using cuvettes which allow for higher signal strength. To optimize signal acquisition we systematically examined the effect of three primary variables on signal strength and measurement noise, something which has not been previously reported in the literature. These variables included PMT voltage, scanning rate and slit width. Raman scattering peak intensities of Milli-Q water, recorded with the different instrument parameter combinations examined, are also reported to facilitate replication of the methods described in this study.

The fluorescence EEMs of GRW (undiluted) and GRW diluted up to 10% of the original strength are shown in Figure 3-1a & b and Figure 3-1c & d respectively. These EEMs were obtained at a scanning rate = 600 nm/min, PMT voltage = 800 V and excitation and emission slit widths = 5 nm each with cuvettes (Figure 3-1a & c) and fibre optic probe (Figure 3-1b & d). The EEM peaks are located at excitation wavelength (Ex) of approximately 330 nm and emission wavelength (Em) between 415-435 nm, in the range reported for HS (Chen *et al.*,

2003; Her *et al.*, 2003; Sierra *et al.*, 2005). In spite of the general spectral similarities in the EEMs obtained with cuvettes and the fibre optic probe in terms of peak positions, there is however significant differences in terms of peak intensities and the signal quality as seen in Figure 3-1a - d.

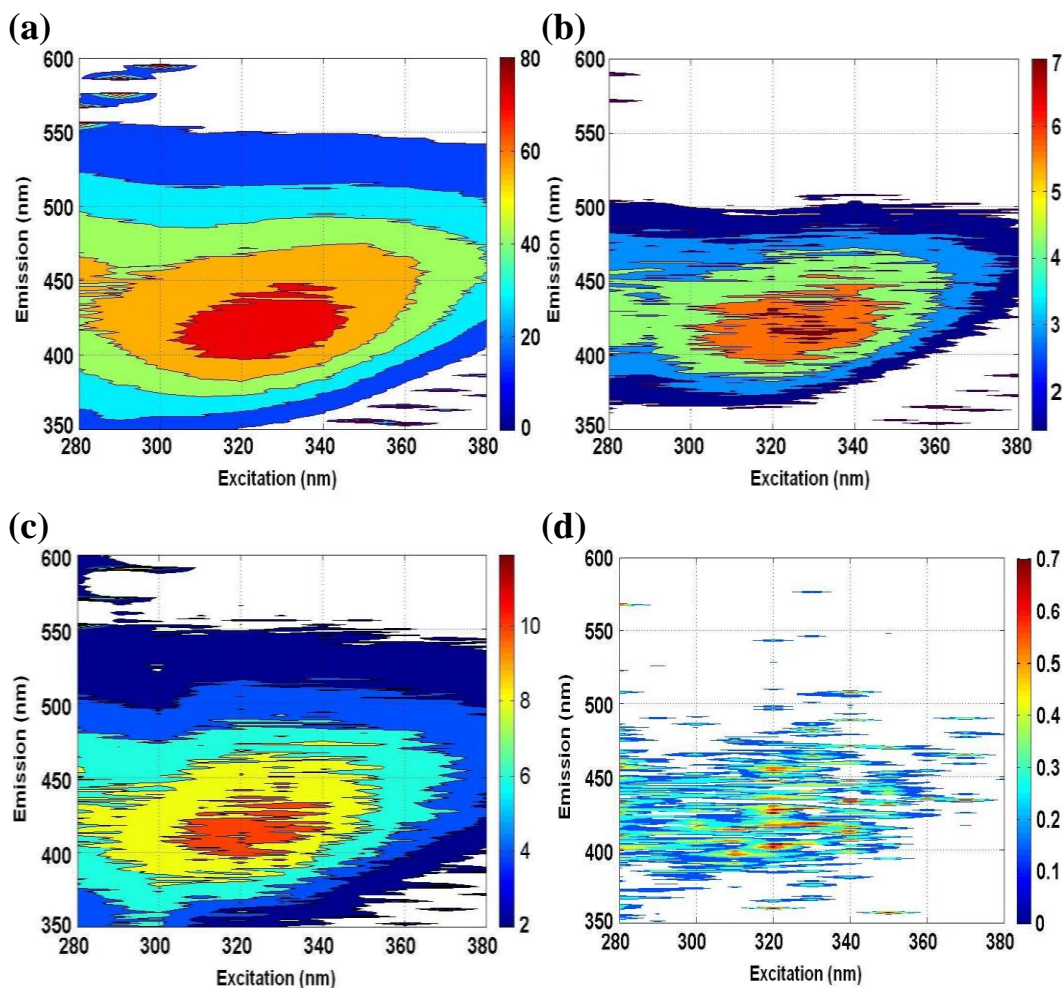


Figure 3-1 Fluorescence EEMs of GRW obtained with (a) cuvettes and (b) fibre optic probe. Fluorescence EEMs of GRW diluted up to 10% of the original strength obtained with (c) cuvettes and (d) fibre optic probe. All measurements were performed at scanning rate of 600 nm/min, PMT voltage = 800 V and excitation and emission slit widths of 5 nm each.

In this study, the weak fluorescence properties of very low concentrations of HS demanded instrument settings that are different from the settings needed for undiluted water samples. Our preliminary studies with PMT voltage in the medium range (600 V), excitation and emission slit widths at 5-10 nm each did not give reproducible spectra, especially for highly diluted samples acquired either with cuvettes or the fibre optic probe. When the fibre optic probe was used to acquire signals from highly diluted samples the problem was more significant; virtually no signal other than noise was observed in some instances (results not shown). This can be explained due to the loss of energy that occurs in the fibre optic assembly when in operation (Lewzey, personal communication). Due to this reason the results presented here were acquired with PMT voltages of 800 V (Figure 3-1) and higher. From the EEM spectra obtained with the given instrument settings in Figure 3-1, it is clear that the fibre optic probe produced noisier spectra. At low concentrations both signal acquisition methods were hampered by significant levels of measurement noise (Figure 3-1c & d; Table 3-1). When the signal was acquired using the fibre optic probe, % error ranged from 11.6 – 34.8%, increasing with increasing dilution up to 10% GRW. Error was reduced when the signal was acquired using cuvettes with a range from 3.6 – 28.2%, increasing with increasing dilutions up to 10% GRW.

Table 3-1 Measurement error observed with the fibre optic probe vs. cuvettes for different dilutions of GRW at PMT = 800 V, scanning rate of 600 nm/min and excitation and emission slit widths of 5 nm each.

% GRW	Probe (PMT = 800 V)			Cuvettes (PMT = 800 V)		
	Peak Intensity (a.u.) at Ex/Em: 330/415 nm	error at 95% confidence	% error	Peak Intensity (a.u.) at Ex/Em: 330/430 nm	error at 95% confidence	% error
10	0.65	0.23	34.85	11.15	3.14	28.18
20	1.22	0.25	20.85	16.81	3.35	19.92
40	2.74	0.44	16.13	31.84	3.87	12.16
60	3.75	0.55	14.73	49.99	3.95	7.90
100	6.40	0.74	11.63	79.03	2.84	3.59
Raman scattering peak intensity at Ex/Em ~ 348 nm /396 nm (a.u.)						
	2.17	± 0.23		13.57	± 0.65	

3.3.1 PMT Voltage

Since the fluorescence EEM spectra of GRW, diluted up to 10% of its original strength, was not clear at PMT voltage = 800 V, the PMT voltage was further increased up to 1000 V in order to obtain a much higher signal strength. The scanning rate, and excitation and emission slit widths were unchanged. This resulted in generally higher EEM peak intensities for highly diluted samples, compared to the spectra acquired at 800 V (Table 3-2). EEM spectra obtained with the fibre optic probe however remained fuzzy with no clearly visible peak as in Figure 3-1d (results not shown).

Table 3-2 Measurement error observed with the fibre optic probe and cuvettes for two different dilution levels of GRW at PMT = 1000V, scanning rate of 600 nm/min and excitation and emission slit widths of 5 nm each.

% GRW	Probe (PMT = 1000 V)			Cuvettes (PMT = 1000 V)		
	Peak Intensity (a.u.) at Ex/Em: 330/415 nm	error at 95% confidence	% error	Peak Intensity (a.u.) at Ex/Em: 330/430 nm	error at 95% confidence	% error
10	0.84	0.98	116.90	84.54	68.28	80.76
20	7.93	4.82	60.74	113.82	64.79	56.92
Raman scattering peak intensity at Ex/Em ~ 348 nm /396 nm (a.u.)						
	25.98	± 3.23		87.95	± 6.40	

Even though higher PMT voltages contributed to higher signal strengths in all samples measured, the error associated with the measurements increased significantly (Table 3-2). This observation is consistent with the theoretical background of fluorescence spectroscopy and other observations made elsewhere (Casado-Terrones *et al.*, 2007). Due to these reasons higher PMT voltages (> 800 V) were deemed to be unfavourable for obtaining reproducible spectra of water samples containing very low DOM concentrations.

3.3.2 Scanning Rate

Another method of reducing signal noise is to reduce the speed at which spectra are acquired. With a high signal averaging time or by keeping smaller intervals between successive data points, the noise associated with measurement can be minimized. This improvement however comes with the cost of increased scanning time. When the scanning rate was reduced from 600 nm/min to 120 nm/min, the measurement noise for low concentration levels was reduced

significantly with the fibre optic probe (Table 3-1 & Table 3-3). The same observation was made when cuvettes were used (results not shown). At a scanning rate of 120 nm/min, it takes almost 24 minutes to acquire a complete fluorescence EEM, with emission scans performed from 350 nm to 600 nm for excitation levels from 280 nm to 380 nm in 10 nm increments. However, at a scanning rate of 600 nm/min, it would only take less than 5 minutes to acquire the same EEM. Long scanning times diminish the likelihood of the fluorescence spectroscopy being used as an on-line monitoring tool. Therefore slower scanning rates were not considered as a viable approach and no further studies were pursued with this parameter.

Table 3-3 Measurement error observed when signal was acquired using a fibre optic probe for two different dilution levels of GRW at PMT = 800 V, scanning rate of 120 nm/min and excitation and emission slit widths of 5 nm each.

% GRW	Peak intensity (a.u.) at Ex/Em: 330/415 nm	error at 95% confidence	% error
10	0.83	0.12	15.02
20	1.53	0.12	7.80
Raman scattering peak intensity at Ex/Em ~ 348 nm /396 nm (a.u.)			
	2.14	± 0.23	

3.3.3 Emission and Excitation Slit Widths

Measurement noise can also be reduced by increasing the excitation energy using larger excitation and emission slit widths. Increasing the slit widths is however constrained to a limit given by the formula (Westerhoff *et al.*, 2001; Lewzey, personal communication):

$$2 \times (Ex \text{ slit width} + Em \text{ slit width}) \leq \text{Stokes Shift} \quad (3.1)$$

For GRW the Stokes shift is approximately 90 nm (Figure 3-2). Therefore the measurement noise levels at different slit width combinations were studied in detail adhering to the above constraint (Equation 3.1). In general, as the excitation and/or emission slit widths were increased the measurement noise decreased in all dilution levels studied. Since the main objective was to reduce measurement noise at low concentration levels only the results relevant to GRW diluted up to 10% of its original strength are presented here (Table 3-4).

Table 3-4 Signal noise at different slit width combinations, obtained with the fibre optic probe and cuvettes on GRW (diluted up to 10%) – calculated at the EEM Peak at Ex: 330 nm; Em: 415 nm.

Emission slit width (nm)	% error at different slit width combinations					
	Excitation slit width (nm)					
	Fibre optic Probe			Cuvettes		
	5	10	20	5	10	20
5	34.9	8.7	6.0	28.2	11.8	10.8
10	8.2	4.5	4.1	13.1	5.0	4.3
20	5.1	4.3	3.8	8.5	3.8	0.4
Raman scattering peak intensity at Ex/Em ~ 348 nm /396 nm (a.u.)						
	12.05	± 0.63		87.21	± 0.83	

From these results (Table 3-4) it is evident that as the excitation and emission slit width is increased, the measurement noise decreases. A similar observation was made, with the same

slit width combinations, when undiluted GRW was analyzed by both cuvette and fibre optic approaches (results not shown). However, slit widths may not be increased without checking the following constraints (Equations 3.2 & 3.3) in order to preserve the resolution of the signal (Creighton et. al., 1997):

$$Ex \text{ slit width} \times 10 \leq Band \text{ Width}_{ex} \quad (3.2)$$

$$Ex \text{ slit width} \times 10 \leq Band \text{ Width}_{em} \quad (3.3)$$

Where $Band \text{ Width}_{ex}$ and $Band \text{ Width}_{em}$ are the band widths of the excitation and emission signals, respectively. The band width of both excitation and emission signals of GRW is approximately 120 nm each (Figure 3-2). According to the above criteria, the maximum slit width levels that can be used without compromising the resolution of the signal are excitation and emission slit widths of 10 nm each. This phenomenon can be further confirmed by examining the spectra of different slit width combinations examined in this study (Figure 3-3). Figure 3-3 illustrates how the signal resolution diminishes as the slit widths are increased beyond 10 nm. The spectral information available in the emission spectra (Figure 3-3; circled regions), in the form of a Raman peak, obtained with emission and excitation slit widths of 10 nm or below, appeared to be absent in the emission spectra obtained with emission and excitation slit widths greater than 10 nm.

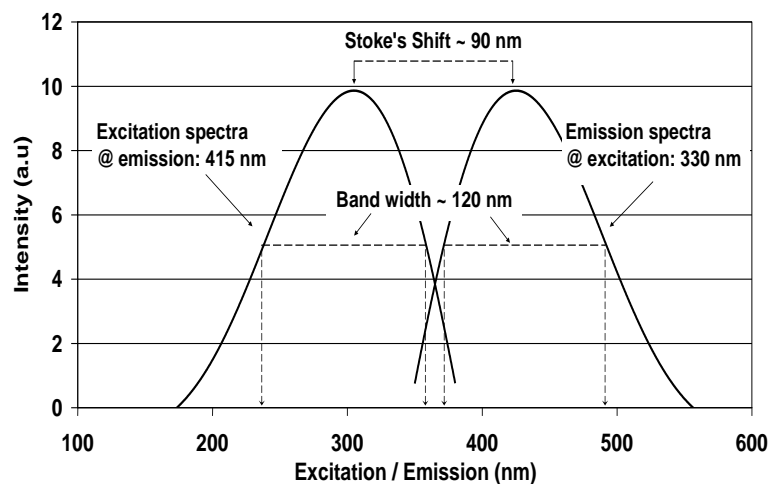


Figure 3-2 Trend lines of excitation and emission spectra of GRW, diluted up to 10% of its original strength, taken at emission = 415 nm and excitation = 330 nm respectively using the fibre optic probe. A scanning rate of 600 nm/min, PMT voltage = 800 V and excitation and emission slit widths of 10 nm each were used.

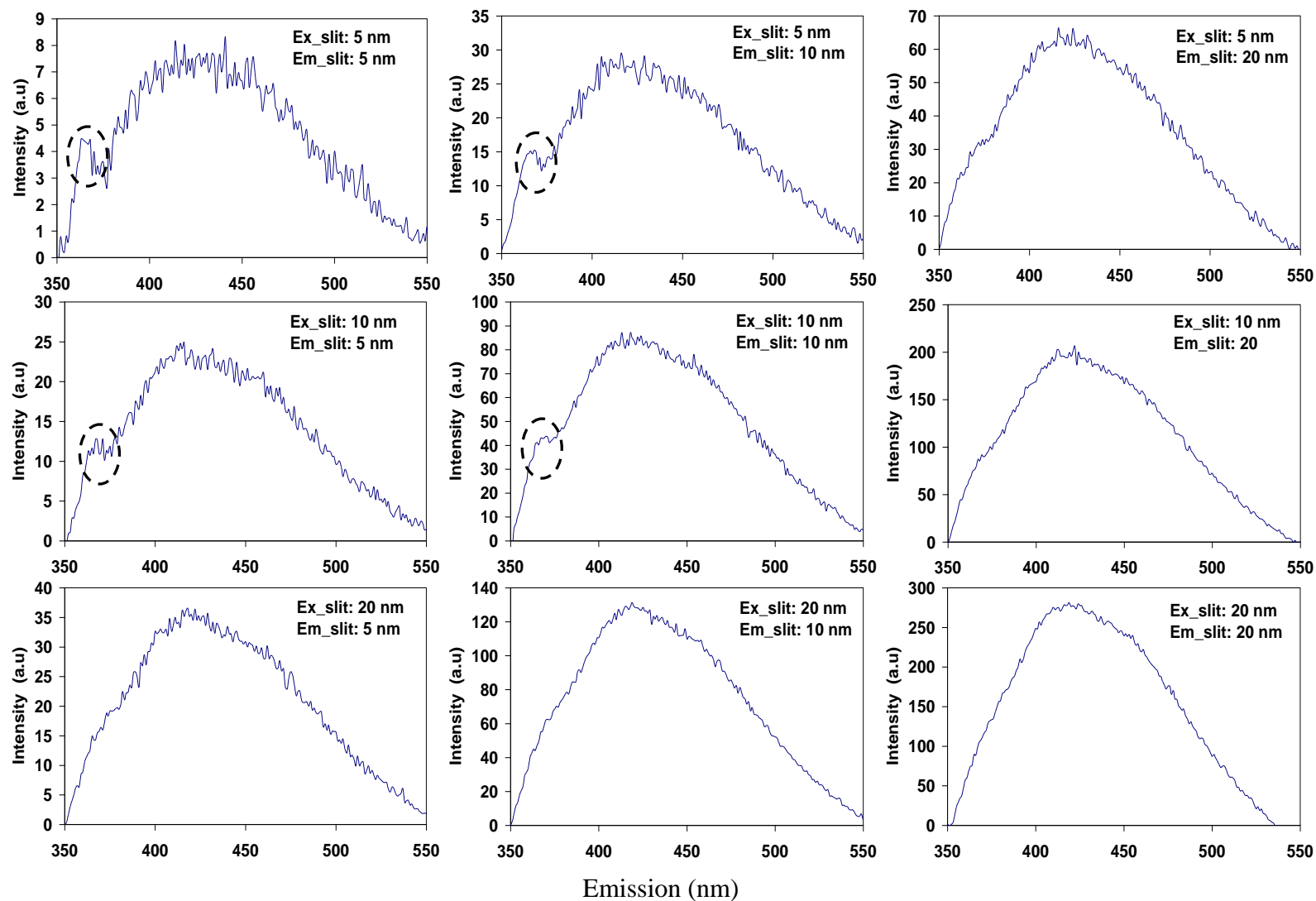


Figure 3-3 Emission spectra of GRW at excitation = 330 nm with different excitation and emission slit width combinations obtained using the fibre optic probe at a scanning rate of 600 nm/min and PMT voltage = 800 V.

Using the optimum instrument parameter settings: PMT = 800 V, scanning rate of 600 nm/min and excitation and emission slit widths of 10 nm each, that were identified in this study, we have obtained reproducible fluorescence EEMs of the residual DOM present in NF permeates that contained DOC about 0.7 mg/L (Peiris *et al.*, 2008). In addition, this parameter combination also allowed us to capture fluorescence EEMs of very low concentration levels of DOM that are present in other water sources such as drinking water and rain water with reasonable reproducibility (i.e. signal noise is < 5 - 10% of the peak EEM intensities measured; results not shown). The manipulation of the instrument parameters demonstrated in this study could therefore serve as a useful guideline for those interested in using fluorescence spectroscopy to analyze very low concentration levels of DOM, without having to pre-concentrate samples using tedious, expensive and time consuming experimental procedures that could contribute to additional experimental error. When implementing these methods, however, instrument biases in different fluorescence spectrofluorometers may need to be accounted for using the Raman scattering peak intensities reported in Table 3-1, Table 3-2 and Table 3-3.

3.4 Conclusions

When acquiring fluorescence signals for very low concentrations of DOM in water, special attention needs to be given to capture reliable signals with good signal resolution. The instrument parameters that are commonly used to produce quality spectra from strongly fluorescing substances proved to be less suitable in dealing with highly diluted HS present in very low DOM concentrations levels. Low PMT voltages (600 V) failed to produce signals

with significant strength when the fibre optic probe was used. Therefore, the PMT voltage was increased up to 800 V to obtain improved signal strength. However, the measurement noise of samples at low concentrations increased significantly when increasing the PMT voltage, and therefore increasing the PMT voltage above 800 V was seen as unfavourable. Nevertheless, even with PMT = 800 V, very poor reproducibility of the fluorescence spectra of highly diluted samples was produced using cuvettes and the fibre optic probe measurements with the later being most affected. Slower scanning rates showed potential for producing good reproducibility. Slower scan rates, below 600 nm/min, were however not seen as a viable option, due to longer scanning times reducing the likelihood of the fluorescence spectroscopy being used as an on-line monitoring tool. As the excitation and/or emission slit widths were increased the measurement noise decreased at all dilution levels studied. This approach was seen as the best method to reduce the measurement noise since it generated the most improvement out of the three approaches studied. Nevertheless, this approach should be practiced with care to avoid diminishing signal resolution. In conclusion, excitation and emission slit widths up to 10 nm are recommended to preserve the spectral information and to obtain reproducible signals with minimum signal noise. The approach demonstrated here shown was successfully applied in the analysis of very low DOM concentration levels and could be useful for researchers interested in using fluorescence spectroscopy to analyze similar concentration levels.

CHAPTER 4

Identification of Humic Acid-like and Fulvic Acid-like Natural Organic Matter in River Water using Fluorescence Spectroscopy

Overview

Identifying the extent of humic acid (HA)-like and fulvic acid (FA)-like natural organic matter (NOM) present in natural water is important to assess disinfection by-products formation and fouling potential during drinking water treatment applications. However, the unique fluorescence properties related to HA-like NOM is masked by the fluorescence signals of the more abundant FA-like NOM. For this reason, it is not possible to accurately characterize HA-like and FA-like NOM components in a single water sample using direct fluorescence EEM analysis. A relatively simple approach is described here that demonstrates the feasibility of using a fluorescence excitation-emission matrix (EEM) approach for identifying HA-like and FA-like NOM fractions in water when used in combination with a series of pH adjustments and filtration steps. It is demonstrated that the fluorescence EEMs of HA-like and FA-like NOM fractions from the river water sample possessed different spectral properties. Fractionation of HA-like and FA-like NOM prior to fluorescence analysis is therefore proposed as a more reasonable approach.

Keywords: Natural organic matter (NOM); fluorescence spectroscopy; fulvic acid-like NOM; humic acid-like NOM

4.1 Introduction

Characterization of natural organic matter (NOM) present in natural waters is essential for understanding and controlling some pertinent water-related problems such as membrane fouling, disinfection by-products formation and undesirable biological growth experienced in water treatment and distribution systems. For example, individual and combined effects that result due to the interplay between different NOM fractions can lead to different membrane fouling behaviour (Jermann *et al.*, 2007). Also, the potential for disinfection by-products formation for different NOM fractions varies considerably (Marhaba *et al.*, 2000). Therefore, knowledge on the major NOM fractions present in water would enable the design and implementation of more efficient treatment processes and control procedures that can mitigate the negative effects of NOM in water treatment and distribution systems.

Major NOM fractions present in natural water are humic substances (HS), protein-like and polysaccharide-like materials, with HS being the most significant component (Thurman, 1985). The aquatic HS are considered to contain largely fulvic acid (FA)-like matter and comparatively smaller amounts of humic acid (HA)-like matter. This is because the lower carboxylic content in HA-like matter lowers its solubility and therefore most natural waters contain 5 to 25 times more FA-like matter than HA-like matter (Thurman, 1985). In membrane-based processes HA-like NOM is known to affect membrane charge more than FA-like NOM and as a result adsorption capacity of HA-like NOM on the membrane surfaces and in the pores of the membranes is higher than for FA-like NOM (Jucker and Clark, 1994). Therefore, increased levels of HA-like matter in natural waters could cause more fouling than FA-like matter (Kulovaara *et al.*, 1999). In addition, the potential for

disinfection by-products formation by aquatic HA-like NOM has also been found to be higher than FA-like NOM (Marhaba *et al.*, 2000).

This study focused on characterizing HA- and FA-like NOM in water as a method of identifying the presence of HA- and FA-like NOM present in water. For this purpose, a fluorescence excitation-emission matrix (EEM) approach was used as it is able to capture specific fluorescence features that correspond to HS and protein-like materials (Coble *et al.*, 1990; Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003; Sierra *et al.*, 2005; Hudson *et al.*, 2008; Henderson *et al.*, 2009). Unlike other available NOM characterization methods (Huber *et al.*, 1992; Her *et al.*, 2003; Gray *et al.*, 2007), fluorescence spectroscopy is able to differentiate the major NOM fractions and is suitable for performing rapid, direct and accurate analysis with high instrumental sensitivity (Coble *et al.*, 1990; Peiris *et al.*, 2008).

The detection of HA-like fluorescence in natural water is more challenging. In most studies the HA-like NOM exhibits a weak fluorescence signal with a peak that appears more or less as a shoulder to the FA-like NOM fluorescence peak (Mobed *et al.*, 1996; Baker, 2001; Sierra *et al.*, 2005; Hudson *et al.*, 2008). Others have observed no significant differences between fluorescence EEMs of aquatic FA-like and HA-like NOM (Alberts and Takács, 2004). These observations suggest that FA- and HA-like fluorescence in most cases overlap, making an accurate identification of HA-like NOM in the presence of FA-like NOM difficult. This difficulty is mainly due to the comparatively weaker fluorescence signals of the less abundant HA-like components which are overshadowed by the stronger fluorescence signals of more abundant FA-like components. Therefore, direct fluorescence EEM

measurement of natural waters may not be suitable for identifying the extent of HA-like NOM in the presence FA-like NOM in natural water.

For these reasons, in this study, fluorescence EEMs of the HA- and FA-like NOM were obtained by fractionating the HA- and FA-like components of NOM from river water using pH changes and filtration. The approach presented here would be useful for those who are interested in a simple approach for characterizing HA- and FA-like matter present in natural waters.

4.2 Methods

4.2.1 Sample Preparation

Grand River water (GRW) (Southwestern Ontario, Canada; collected on July 05, 2009) was filtered through a 0.45 μm GN-6 Metricel® type 47 mm diameter disc filter (Pall Corporation, Ann Arbor, MI; Lot #: 73056) to remove particulate matter. The pH of this filtered GRW was ~ 8.2 . The pH of the GRW ($\sim 1\text{L}$) was then adjusted below $\text{pH} = 1$ by adding high purity hydrochloric acid (1 mol/L HCl, 99.999%). At this pH, it is expected that HA-like NOM in the water sample should precipitate (Aiken, 1985; Ghabbour *et al.*, 1994). The pH adjusted water sample was filtered again through a 0.45 μm filter to remove the precipitate. The pH of this filtered water sample was adjusted back to the initial pH level of $\text{pH} \sim 8.2$ using high purity NaOH (1 mol/L, 99.998 %). This filtered water sample can be expected to contain the FA-like fraction of the HS present in GRW (Aiken, 1985; Ghabbour *et al.*, 1994). The precipitate collected on the filter membrane was carefully scrapped from

the surface of the membrane and dissolved in 50 ml of Milli-Q (Millipore) water (resistivity = 18.2 M Ω cm). The pH of this dissolved precipitate was adjusted back to pH ~8.2. During all pH adjustments the addition of HCl acid and NaOH were maintained to a minimum (only a few drops) to minimize the dilution of water samples. The sample preparation method is illustrated in Figure 4-1. As shown, three samples resulted that included: (i) FA-like NOM solution at pH ~ 8.2; (ii) HA-like NOM extract solution at pH < 1 and (iii) HA-like NOM extract solution at pH ~ 8.2. These samples were analyzed by fluorescence spectroscopy.

In addition, fluorescence EEM of Aldrich humic acid (AHA) obtained from Sigma Aldrich Inc., (St. Louis, MO; Lot #: 100K0219, Alginic acid, sodium salt) was used to validate the existence of HA-like matter in the HA-like NOM extract of GRW.

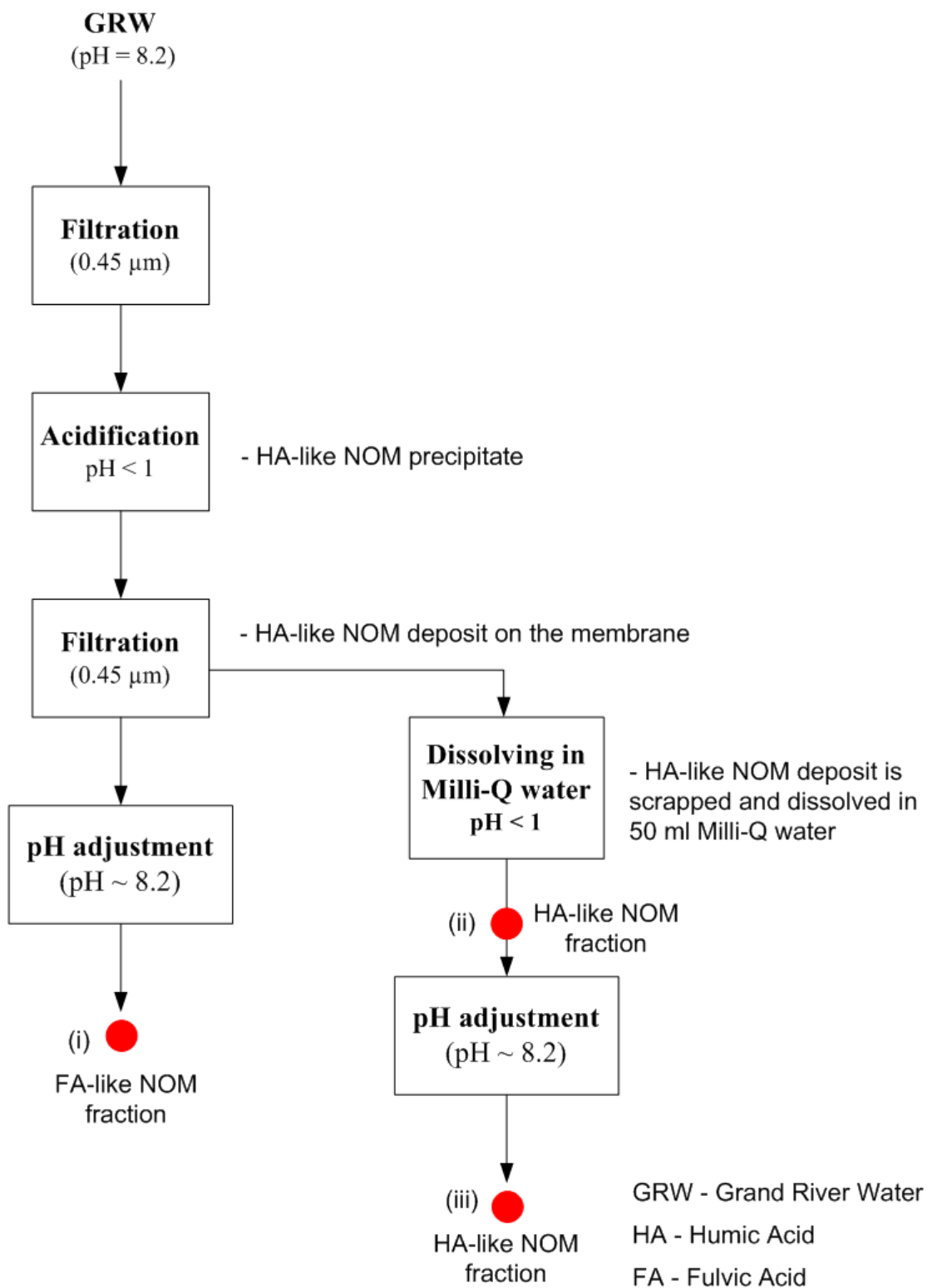


Figure 4-1 Procedure for the fractionation of HA- and FA-like NOM fractions from GRW.

● - sampling points for fluorescence analysis.

4.2.2 Fluorescence Analysis

The fluorescence EEMs were recorded using a Varian Cary Eclipse Fluorescence spectrofluorometer (Palo Alto, CA) collecting 301 individual emission intensities (Em: 300 - 600 nm) at sequential 10 nm increments of excitation wavelengths between 250 and 380 nm. Instrument parameters: photomultiplier tube (PMT) voltage = 800 V, scan rate = 600 nm/min and excitation/emission slit width = 10 nm each were maintained during the fluorescence signal acquisition. A more detailed description of the fluorescence analysis procedure and the selection of the spectrofluorometer parameter settings used in this study for obtaining reproducible fluorescence signals, especially for low NOM concentrations levels will be found in Peiris *et al.* (2008 and 2009). Corrections for inner filter effects were not applied as inner filtering effects are not expected to be significant at the low concentration levels examined in this study (Hudson *et al.*, 2008). To eliminate water Raman scattering and to reduce other background noise, fluorescence spectra for Milli-Q water, obtained under the same conditions, were subtracted from all spectra. During the course of the fluorescence analyses, the change in Raman scattering peak intensities recorded for Milli-Q water at Ex/Em ~ 348 nm/396 nm was less than 1%, confirming that there were no significant fluctuations in the performance of the spectrofluorometer lamp or other hardware. The water samples were maintained at room temperature (~ 25 °C) during the analyses.

4.3 Results and Discussion

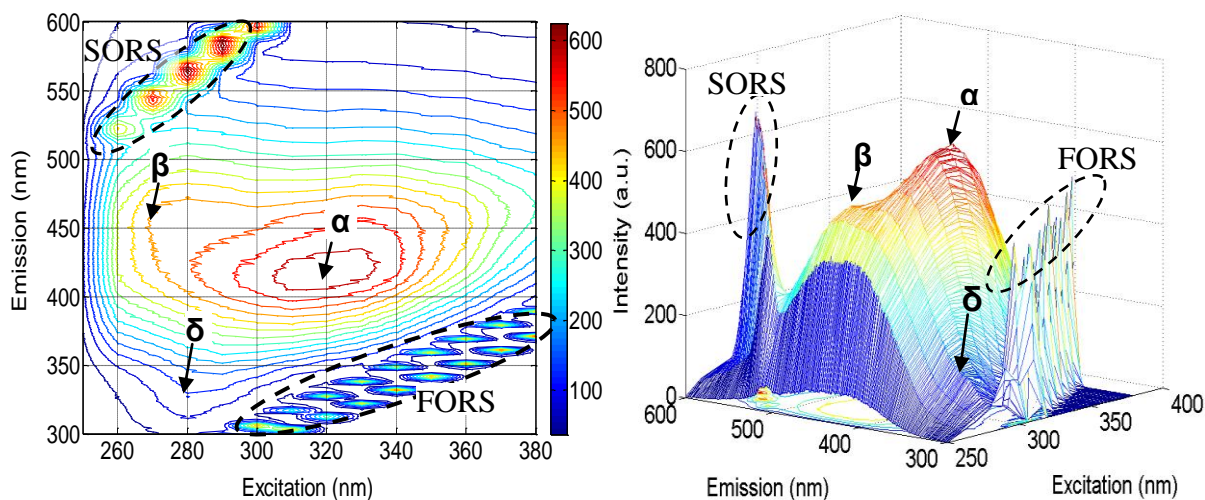


Figure 4-2 Fluorescence EEM for GRW (left) and 3D view of the same EEM (right). First order Rayleigh scattering (FORS) and second order Rayleigh scattering (SORS) regions are indicated using dashed-circles.

4.3.1 Typical Spectral Features of GRW

The fluorescence EEM of GRW used in this study shows a peak (α) at the excitation wavelength (Ex) and the emission wavelength (Em) combination: Ex/Em = 320 nm/415 nm (Figure 4-2), which corresponds to the range reported for FA-like NOM in water (Coble *et al.*, 1990; Sierra *et al.*, 2005). Another secondary peak (β) which also corresponds to HS (Sierra *et al.*, 2005; Peiris *et al.*, 2008) appears to be present in the form of a shoulder to the main peak (α) around Ex/Em = 270 nm/460 nm (Figure 4-2). The presence of HS in GRW that is associated with these spectral features was also independently confirmed by the Liquid Chromatography - Organic Carbon Detection (LC-OCD) chromatograph analysis in a previous study (Peiris *et al.*, 2008). The HS in GRW can be expected to comprise predominantly of FA-like matter compared to HA-like matter as reported in other natural waters (Huck, 1999; Sierra *et al.*, 2005). The deviations of the fluorescence EEM contours seen in the region indicated by δ (Ex/Em: 280 nm/330 nm) is due to the presence of protein-

like NOM in the water; a fluorescence EEM peak around the same region (δ) has been previously observed for protein-like substances (Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003). The existence of this protein region (δ) however does not appear to be a significant fluorescence peak due to the very low concentration levels of the protein-like substances present in the GRW. The light scattering regions (first order Rayleigh scattering region and second order Rayleigh scattering region) observed in the fluorescence EEM are also important areas that provide information related to the particulate/colloidal matter present in water (Peiris *et al.*, 2010).

4.3.2 Fluorescence EEMs of FA-like and HA-like NOM in GRW

The fluorescence EEM of FA-like NOM (Figure 4-3a) demonstrates similar fluorescence spectral characteristics to that of GRW. This indicates that GRW contained predominantly FA-like NOM as expected in river water. However, Figure 4-3a also demonstrates a shift (about 10 nm) in the peak (α) towards the higher excitation wavelength as illustrated by peak (α'). The fluorescence EEM of HA-like NOM extract at pH < 1 (Figure 4-3b), on the other hand, does not seem to contain fluorescence spectra related to HS-like matter that is seen in Figure 4-2. This can be explained by the poor solubility of HA at pH < 1 (Aiken, 1985) resulting in little or weak fluorescence.

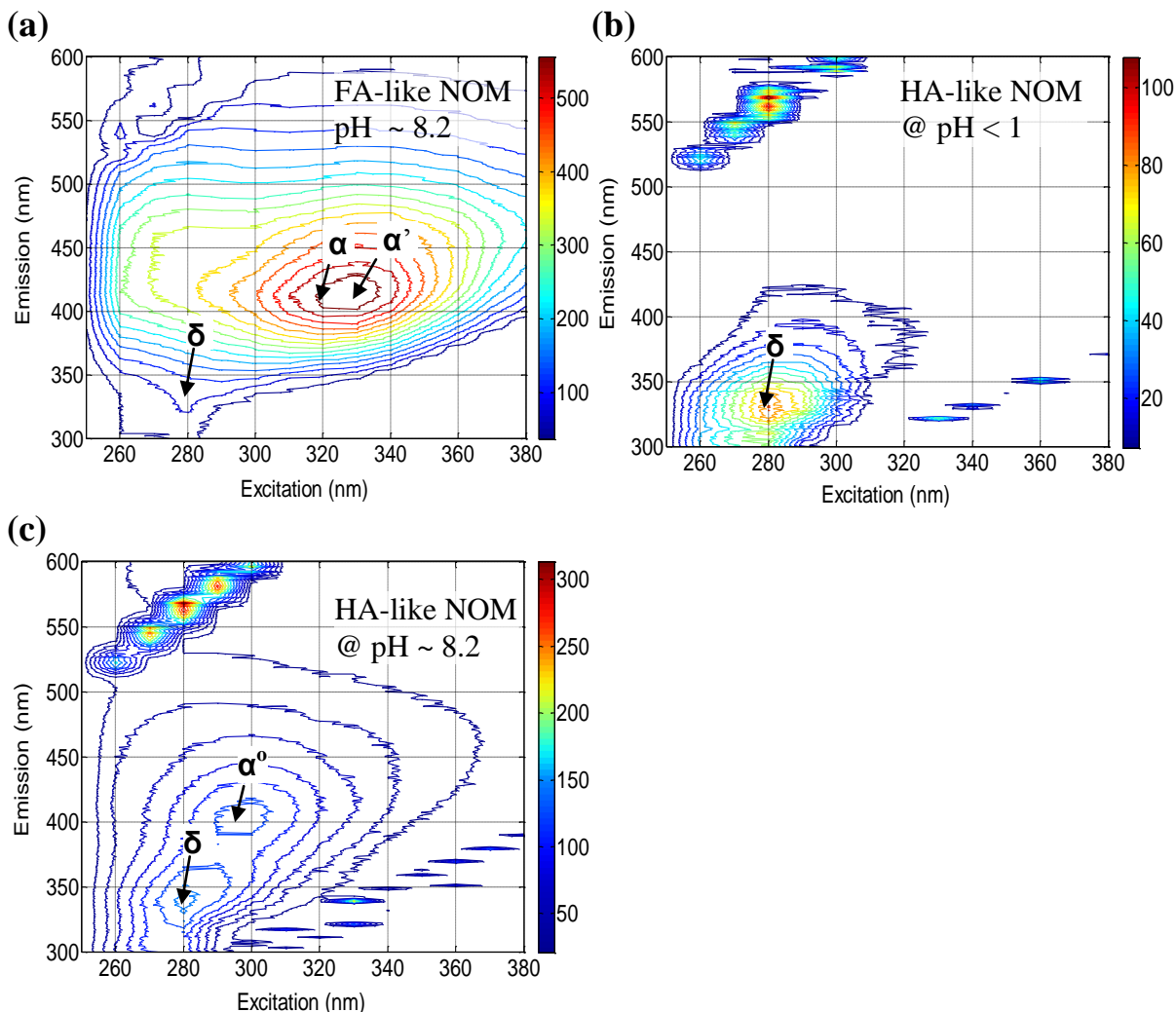


Figure 4-3 Fluorescence EEMs of (a) FA-like NOM – obtained at pH ~ 8.2; (b) HA-like NOM extract at pH < 1 and (c) HA-like NOM extract at pH ~ 8.2. The pH adjustments of the water samples were made using HCl and NaOH.

When the pH of the HA-like NOM extract was increased to the normal pH level of GRW (pH ~ 8.2), a new fluorescence peak (α^0) appeared at Ex/Em = 296 nm/400 nm (Figure 4-3c). This peak fell within the general fluorescence region related to the HS-like NOM in GRW as demonstrated in Figure 4-2 (i.e. between peak (α) and (β)). It also has a closer resemblance to the fluorescence EEM of Aldrich humic acid (AHA) obtained under the same conditions (Figure 4-4) with respect to the excitation wavelength at which the EEM peak is situated.

The fluorescence EEM peak of HA-like NOM extract of GRW is situated at $\text{Ex} = 296 \text{ nm}$ and that of AHA at around $\text{Ex} \sim 295 \text{ nm}$ (Figure 4-3c and Figure 4-4). The difference in the EEM peak position of AHA and HA-like NOM in GRW on the emission scale is expected as similar differences were also observed in other studies in which AHA and HA extracts of river water were used (Parlanti *et al.*, 2000; Sierra *et al.*, 2005). The existence of peak (α^0) at $\text{pH} \sim 8.2$ but not at $\text{pH} < 1$ can be related to the HA-like NOM being completely re-dissolved at higher pH levels and hence its fluorescence properties.

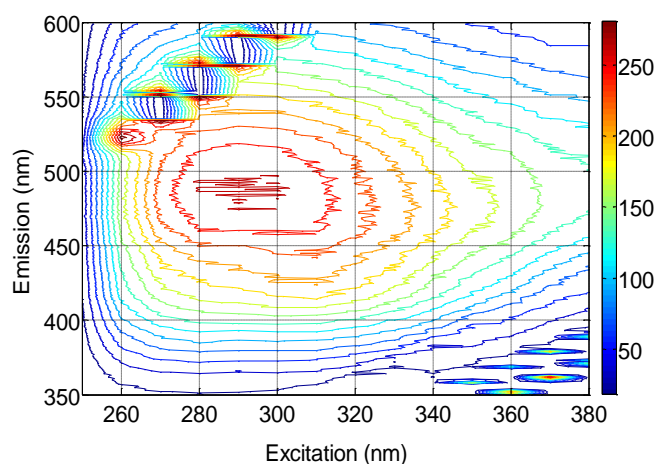


Figure 4-4 Fluorescence EEM of Aldrich humic acid obtained at $\text{pH} = 8.2$.

In addition, the presence of a clear protein-like fluorescence EEM peak in the HA-like NOM fraction indicates that some of the protein-like NOM present in GRW was also present in this HA-like NOM extract (Figure 4-2, Figure 4-3b and Figure 4-3c). The spectral properties related to the protein-like NOM is also present in the fluorescence EEM of the FA-like NOM fraction of GRW (δ -region in Figure 4-3a). Therefore, it appears that protein-like NOM in GRW is soluble at lower pH conditions that were used during the fractionation of HS. Nevertheless, since the fluorescence EEM peak related to protein like NOM is situated

reasonably outside the fluorescence EEM region of HS, the presence of protein-like NOM in both HA-like and FA-like fractions did not impede the detection of HA-like and FA-like fluorescence signals.

4.3.3 Significance of this Approach and Potential for Future Research

This study demonstrates that fluorescence signals related to HA-like NOM extracts of GRW are located within the general fluorescence region related to HS-like NOM. Therefore, if unfractionated natural water were analysed using fluorescence EEM, one would not expect to be able to clearly identify the existence of HA-like NOM as the unique fluorescence properties related to HA-like NOM is overshadowed by the fluorescence signals of the more abundant FA-like NOM. This has been the case in many fluorescence-based analyses as described in the introduction. Since the predominant component of HS present in natural water is FA-like matter, fluorescence spectral properties of natural water is largely associated with fluorescence properties of the more abundant FA-like NOM. Due to these reasons, one cannot make an accurate characterization of the HA-like NOM present in natural water samples by direct fluorescence analysis. In this context, fractionation of natural water samples prior to fluorescence analysis, as proposed in this study, could provide more information on the content of HA-like and FA-like NOM in natural water. The approach presented here is relatively simple to perform and hence it could be useful for those with an interest in the effect of HA-like and FA-like NOM in water treatment processes and distribution systems.

4.4 Conclusions

In this study, we have performed fluorescence EEM-based characterization of HA-like and FA-like NOM in GRW by fractionation using a combination of pH adjustments and filtration steps. The HA-like NOM fraction of GRW exhibited a fluorescence EEM peak that lies within the general fluorescence region related to HS-like NOM. This highlights the challenges of performing direct fluorescence-based analysis when interested in detecting HA-like NOM in natural water as the unique fluorescence properties related to HA-like NOM are overshadowed by the fluorescence signals of the more abundant FA-like NOM. A simple pH-based separation of HA- and FA-like NOM fractions is proposed followed by fluorescence EEM analysis that could be helpful for obtaining more accurate qualitative information on the HA- and FA-like NOM content in natural water.

CHAPTER 5

Development of a Fluorescence-based Soft Sensor Approach to Differentiate and Quantify Natural Organic Matter in Water

Overview

Understanding the characteristics and quantification of natural organic matter (NOM) is important for the development and operation of efficient drinking water treatment and distribution systems. Although various characterization methods of NOM have been reported, the rapid quantification of different fractions of NOM in natural water remains as a major challenge. Fluorescence spectroscopy has been widely used to characterize NOM in water and is well known as a rapid and sensitive characterization method. Most reported fluorescence-based techniques however have focused on qualitative characterization of NOM but not on quantification of the different NOM fractions in water. This study focused on the development of a fluorescence-based soft sensor (FSS) approach as a method of quantification and differentiation of the two major NOM fractions in water: humic acid (HA)- and fulvic acid (FA)-like matter. In this approach, fluorescence intensity data for these two different NOM fractions were acquired using a fluorescence emission-excitation matrix (EEM) method. These spectral data were then used as the basis for quantification of the NOM fractions using a previously developed calibration model. The calibration model used here was developed by correlating fluorescence spectral data with independent concentration measurements based on dissolved organic carbon (DOC) measurements

determined independently using Partial Least Squares (PLS) regression. The proposed FSS approach was suitable for quantifying and differentiating HA- and FA-like NOM with good accuracy showing its potential for quantification of NOM in complex water matrices of natural water.

Keywords: fluorescence spectroscopy; partial least squares regression; natural organic matter; soft sensor

5.1 Introduction

The presence of natural organic matter (NOM) in natural waters creates many challenges for drinking water treatment and distribution operations such as increasing the distribution of organic micro-pollutants, promoting undesirable biological growth in distribution systems (Christy *et al.*, 2000), the formation of disinfection by-products (DBP) during disinfection/oxidation of water (Marhaba *et al.*, 1998) and membrane fouling. Since the extent of these negative effects is dependent on different NOM fractions such as protein, humic acid (HA)-, fulvic acid (FA)- and polysaccharide-like substances, understanding their characteristics and developing tools to quantify the fractions are required for the development and the implementation of monitoring and control strategies. Even though, a number of characterization methods for NOM have been reported (Huber *et al.*, 1992; Her *et al.*, 2003), the reliable quantification of different fractions of NOM in natural water using simple methods remains as a major challenge.

The heterogeneous and complex nature of NOM presents some unique challenges when using established characterization techniques (Christy *et al.*, 2000). Most of the characterization methods (Huber *et al.*, 1992; Her *et al.*, 2003) require concentration and/or fractionation of NOM before analysis due to their limited sensitivity to the low NOM concentration levels found in natural water (Peiris *et al.*, 2008). Aside from being unsuitable for on-line applications, the existing characterization methods are time consuming, expensive, complex and are based on difficult analyses that require operator expertise (Her *et al.*, 2003). In this context, fluorescence spectroscopy has received significant attention for the characterization of NOM and shows good potential for on-line monitoring, as minimal sample pre-treatment and preparation are required, high instrumental sensitivity is available and the technique is non-destructive in nature (Coble *et al.*, 1990; Peiris *et al.*, 2008 and 2009). In this study, the fluorescence excitation-emission matrix (EEM) approach was used as it is able to capture specific fluorescence features that correspond to NOM in water (Coble *et al.*, 1990; Chen *et al.*, 2003; Her *et al.*, 2003; Sierra *et al.*, 2005).

The importance of analyzing all the fluorescence intensities captured at different excitation- (Ex) and emission-wavelength (Em) combinations as opposed to analyzing only major fluorescence peak positions (peak-picking method) has been emphasized in many studies for capturing the heterogeneity of the different NOM fractions in water (Chen *et al.*, 2003; Stedmon *et al.*, 2003). Also, analyses of the full fluorescence EEM as opposed to only the main peak positions has been observed to provide better sensitivity in terms of identifying the changes that occur in NOM fractions in water (Peiris *et al.*, 2010).

Some reported characterization techniques, such as Fluorescence Regional Integration (FRI) (Chen *et al.*, 2003) and the PARAFAC approach (Stedmon *et al.*, 2003), take into account all the data points in the full fluorescence EEM. These techniques also have certain limitations, for example the FRI method is based on the ad-hoc division of EEMs into smaller regions to capture all the unique spectral details of different NOM fractions for increased accuracy. However, there is no systematic method to determine the sub-regions to be used for these calculations. PARAFAC in contrast uses all pairs of Ex/Em data in an EEM. The limitation with the PARAFAC approach is that it is not suitable for the quantitative analyses of different NOM fractions. In this context, considering every data point (i.e. smallest sub-region) in the full fluorescence EEM for analysis have the potential to accurately capture the heterogeneity of the different NOM fractions in natural water. The treatment of all the intensities corresponding to individual combinations of emission-excitation wavelengths as independent variables will be referred in this study as the fluorescence-based soft sensor (FSS) method.

In this study, we report the development of a FSS approach which proved suitable for predicting and quantifying two types of HA- and FA-like NOM that are present in natural water. Grand River water (GRW), (Southwestern Ontario, Canada) was selected to represent FA-like NOM in water as it contains predominantly FA-like NOM. This was confirmed by separate studies: (1) that included the GRW samples used in this study (Peiris *et al.*, 2010) and (2) pH-based separation of HA- and FA-like fractions in GRW following the work by Aiken (1985) as demonstrated in Chapter 4. GRW was also chosen as a source of FA-like NOM within a more representative complex natural water matrix. Aldrich humic acid (AHA)

was chosen to represent HA-like NOM in natural water. Even though AHA is largely a soil-based substance (Malcolm and MacCarthy, 1986), it is reasonable to assume that HA-like NOM, present in natural water, takes its origin from the soil closer to the water sources. However, the fluorescence EEM of AHA demonstrates closer resemblances to fluorescence EEMs of natural water based HA-like matter (Parlanti *et al.*, 2000; Sierra *et al.*, 2005). This was also verified in the fluorescence EEM of the HA fraction of GRW obtained following the work by Aiken (1985) (Chapter 4). Furthermore, AHA has been widely used as a surrogate for HA-like NOM in many water research studies (Jeanneau *et al.*, 2007; Jermann *et al.*, 2009). The rationale in using the proposed technique for mixtures of GRW and AHA is that it is possible to test the prediction accuracy of the method for different specified concentrations of AHA and GRW *a priori*.

The ability of this approach to differentiate and quantify HA- and FA-like NOM within complex water matrices would be helpful for researchers who study the impact of these NOM fractions on membrane fouling and DBP formation. It is proposed that this approach can be extended to other NOM fractions in natural water (work in progress). The soft sensor was designed by correlating fluorescence EEM data of NOM samples with corresponding dissolved organic carbon (DOC) measurements determined independently. Partial Least Squares (PLS) regression, a multivariate statistical technique, was used for developing a calibration model which correlates the intensity data of full fluorescence EEMs of water samples containing different concentration levels of two types of NOM with the corresponding concentration values measured in terms of DOC (mg/L). The choice of PLS

regression is justified by its ability to handle the high collinearity present among the fluorescence EEM intensities (Elshereef *et al.*, 2006).

5.2 Materials and Methods

The two sources of NOM used for sample preparation in this study were GRW (collected on October 15, 2007) and AHA obtained from Sigma Aldrich Inc., St. Louis, MO (batch #: 04013JD, humic acid-sodium salt).

5.2.1 Sample Preparation.

A stock solution of AHA, about 2 g/L, was made by dissolving 2 g of AHA in Milli-Q (Millipore) water (resistivity = 18.2 M Ω cm). The AHA stock solution was subsequently diluted in Milli-Q water to prepare diluted concentrations of AHA solution as needed. All the samples withdrawn from the stock solution to prepare diluted AHA samples were filtered using 0.45 μ m Nylon Acrodisc type syringe filters to remove particulate matter. For the same reason, GRW (about 4 L) was also filtered through 0.45 μ m GN-6 Metrical® 47 mm disc filters. The DOC content (mg/L) of GRW and AHA solutions was also recorded using an OI-Analytical TOC analyzer (Model 1010, College Station, TX) with a wet-oxidation method as described in Standard Methods (2005) 5310D. Different amounts of GRW and AHA were then used to prepare calibration and validation sample sets as explained below.

5.2.2 Calibration Sample Sets

Four sets of sample mixtures, with 10 samples in each set, were made with each set containing different amounts of AHA and a fixed amount of GRW as indicated in Table 5-1. Different concentrations of GRW for a given set were obtained by using different dilutions of GRW (10%, 25%, 50% and 75% of the original strength or 0.65, 1.63, 3.25 and 4.88 mg-DOC/L, respectively). In this way, the total DOC (i.e. DOC of AHA and GRW) in the overall calibration sample set was varied from ~ 1 – 10 mg-DOC/L to cover typical DOC reported in natural surface waters (Thurman, 1985) . The DOC content of GRW and AHA in each sample was estimated based on the DOC content (mg-DOC/L) of the original GRW and AHA solutions used for sample preparations. To examine the reproducibility of the sample preparation method, two sample mixtures from each set were randomly selected to prepare sample triplicates.

5.2.3 Validation Sample Sets

In addition to the calibration set, another four sets of samples (12-14 samples in each set and 53 samples in total) were prepared, following the same procedure as above, with each set having one of the four dilution levels of GRW that was used in the calibration samples (Table 5-1). The AHA concentration of each sample was different from the AHA concentration of the corresponding calibration sample set.

Table 5-1 Concentrations of AHA and GRW in the calibration and validation sample sets.

Sample #		Calibration set										Validation set									
		1	2	3	4	5	6	7	8	9	10	(mg-DOC/L)									
Sample set 1	AHA	0.5	1.0*	1.5	2.0	2.5	3.0	3.5	4.0*	4.5	5.0	0.6	0.8	1.8	3.4	4.4	5.6	6.4	7.4	8.4	
	GRW					0.65						9.6	11.2	14.0		0.65					
Sample set 2	AHA	0.5	1.0	1.5	2.0*	2.5	3.0	3.5	4.0	4.5	5.0*	0.4	0.8	1.6	2.6	3.4	4.8	6.2	7.2	8.2	
	GRW					1.63						9.2	11.0	12.8	13.6	14.6		1.63			
Sample set 3	AHA	0.5*	1.0	1.5	2.0	2.5	3.0*	3.5	4.0	4.5	5.0	0.6	1.4	2.4	3.4	4.6	5.6	7.2	8.2	9.4	
	GRW					3.25						10.5	11.4	12.6	14.4		3.25				
Sample set 4	AHA	0.5	1.0	1.5*	2.0	2.5	3.0	3.5*	4.0	4.5	5.0	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	
	GRW					4.88						9.5	10.5	11.0	12.5	14.1		4.88			

* - denotes samples made in triplicates to examine reproducibility of the sample preparation method.

5.2.4 Fluorescence Analysis

The fluorescence EEMs of each sample was recorded, using a Varian Cary Eclipse Fluorescence Spectrofluorometer (Palo Alto, CA) by scanning 251 individual emission spectra (350 - 600 nm) at sequential 10 nm increments of excitation wavelengths between 250 and 380 nm. A detailed description of the fluorescence analysis procedure and the selection of the spectrofluorometer parameter settings used for obtaining reproducible fluorescence signals, especially for low NOM concentrations levels, is found in Peiris *et al.* (2008 and 2009). Corrections for inner filter effects were not applied as inner filtering effects were not expected to be significant at the low concentration levels within the calibration range examined (Hudson *et al.*, 2008). To eliminate water Raman scattering and to reduce other background noise, fluorescence spectra for Milli-Q water, obtained under the same conditions, were subtracted from all spectra. During the course of the fluorescence analyses, the Raman scattering peak intensity recorded for Milli-Q water at Ex/Em ~ 348 nm/396 nm was less than 1%, confirming that there were no significant fluctuations in the performance of the spectrofluorometer lamp or other hardware. The pH of all the samples were adjusted to ~ 8.2 to match the pH of GRW using high purity 0.1 M HCl acid (99.999%) and 0.1 M NaOH (99.998 %) before the fluorescence analyses.

5.2.5 Soft-sensor Development

The fluorescence EEM of each sample contained 3514 excitation and emission coordinate points. The fluorescence intensity values corresponding to all 3514 coordinate points (spectral variables) of each EEM were rearranged according to the fluorescence EEM data rearrangement procedure described in Peiris *et al.* (2010) and Elshereef *et al.* (2006). This

procedure generated a 40×3514 data matrix (X_c) from the calibration sample set and a 53×3514 data matrix (X_v) from the validation sample set. Each row of these data matrices corresponded to one sample and the intensity values of the corresponding EEM were arranged over 3514 columns. The measured DOC values of AHA and GRW for each sample of the calibration set was also arranged into a 40×2 data matrix (Y_c) with each row containing the DOC content of AHA and GRW.

The X_c and Y_c data matrices were then used as the predictor data set and the response data set respectively to develop a calibration model using PLS regression. PLS regression analysis is a well-known technique for developing calibration models which can be used to correlate between multiple input variables and multiple response variables. In the current study, the input variables are the fluorescence spectral variables in X_c and the response variables are the NOM concentrations measurements in Y_c . PLS is also able to handle problems associated with noise and collinearity present among these fluorescence spectral variables. Essentially, PLS regression analysis extracts a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and mathematically represented by linear combinations of original predictor variables. These linear combinations, referred to as latent variables (LVs), are calculated to account for as much of the manifest covariance existing between predictor data (X_c) set and the response data set (Y_c) (Qin *et al.*, 1992). Both X_c and Y_c data sets were mean-centered before PLS regression analysis. The optimum number of LVs and the goodness of predictions (Q^2) were determined by a cross-validation algorithm (Wold *et al.*, 2001). The validation data set (X_v) was subsequently used to test the prediction accuracy of the developed calibration model. All computations were performed

using the PLS Toolbox 3.5 (Eigenvector Research, Inc., Manson, WA) within the MATLAB 7.3.0 computational environment (MathWorks, Natick, MA).

5.3 Results and Discussion

The fluorescence EEMs of AHA and GRW are shown in Figure 5-1a and b, respectively. The EEM peak position of AHA (Figure 5-1a) is located at Ex/Em \sim 290 nm/490 nm. The EEM peak (α) of GRW, located at Ex/Em \sim 320 nm/415 nm as shown in Figure 5-1b, is within the range reported for FA-like NOM (Coble *et al.*, 1990; Sierra *et al.*, 2005). In addition to peak (α), a second peak (β) which also corresponds to humic substances (Sierra *et al.*, 2005; Peiris *et al.*, 2008), appears to be present in the form of a shoulder of the peak (α) at Ex/Em \sim 270nm/460nm. Figure 5-1c shows fluorescence EEM when AHA and GRW are mixed together. The concentrations of the AHA and GRW in this mixture are about 2 mg-DOC/L and 1.6 mg-DOC/L, respectively. The mixture also has the two peaks observed for GRW with peak (γ) located at Ex/Em \sim 310 nm/420 nm. The second peak (δ) appears as a shoulder at Ex/Em \sim 280nm/455nm. In addition to these peaks, the EEM of the mixture has spectral characteristics that are distinctly different from those seen on the EEMs of the individual components. These differences are primarily due to the presence of AHA and GRW in the mixture. As the concentrations of AHA and GRW are varied in the mixtures, the resulting fluorescence EEMs were also observed to vary (results not shown). When comparing the fluorescence intensity values of both peak (α) and peak (β) of the samples made in triplicate for randomly selected sample mixtures (Table 5-1), the standard deviation (n=3)

corresponding to mixtures did not exceed 5% of the average peak intensity in all cases, confirming the reproducibility of the sample preparation method.

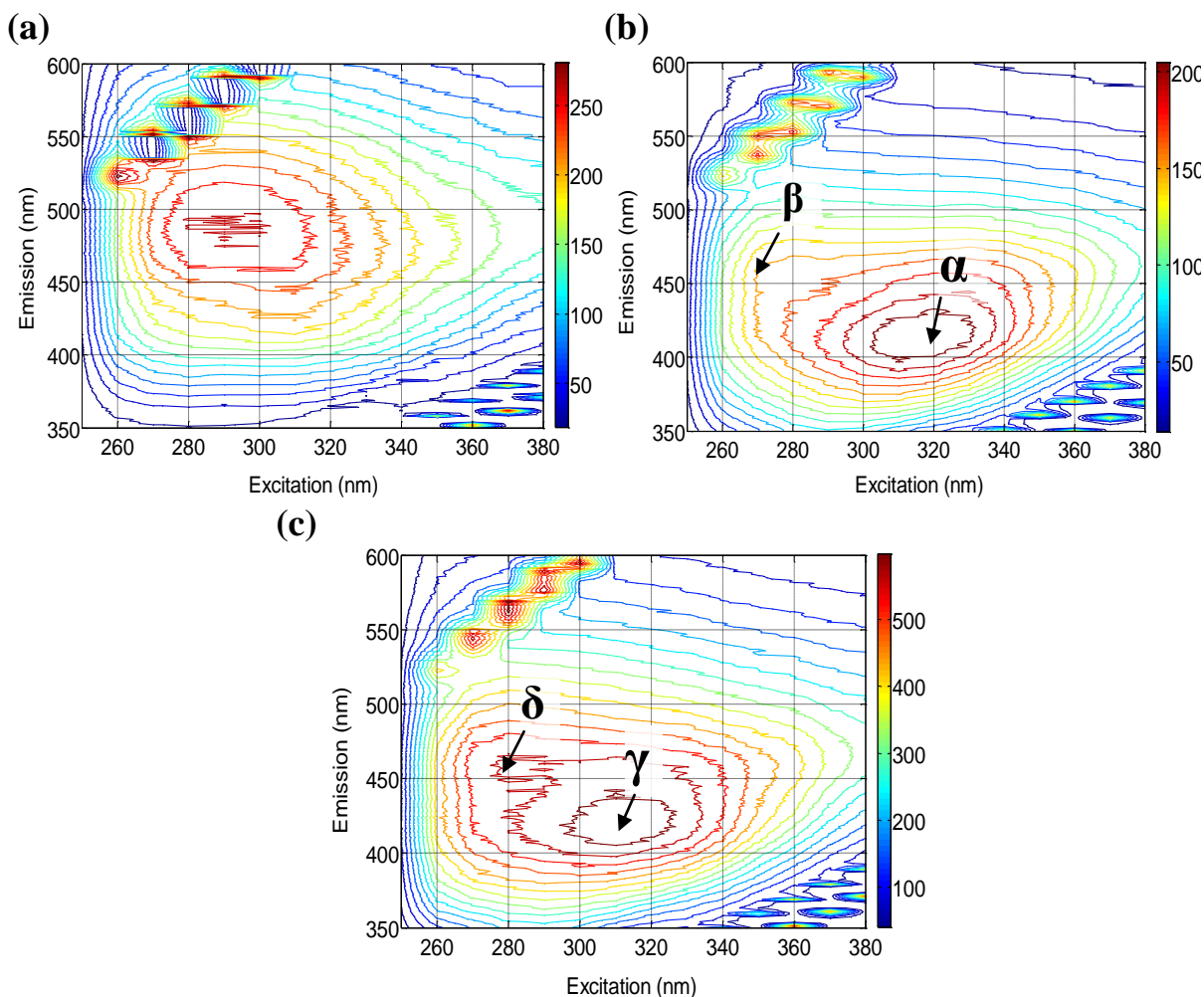


Figure 5-1 Fluorescence EEM of (a) AHA (b) GRW and (c) a mixture of AHA and GRW.

5.3.1 Calibration Model Development

The spectral differences seen in the fluorescence EEMs of the mixtures at different concentrations of AHA and GRW were used to quantify and differentiate the amounts of AHA and GRW in the sample mixtures. By correlating the intensity reading of 3514 spectral variables in the data matrix X_c with the corresponding concentrations of AHA and GRW in

the data matrix Y_c , the PLS regression analysis resulted in a regression model with a relatively small number of statistically significant LVs. The optimum number of LVs statistically significant for accurate predictions was determined as three using the leave-one-out cross-validation method (Qin *et al.*, 1992; Wold *et al.*, 2001). The resulting calibration model with these three LVs demonstrated a goodness of prediction by cross validation of about 98%, capturing 97.7% of the variance in the X_c matrix and 98.3% variance in the Y_c data matrix. The rest of the variability is assumed to be due to experimental error and instrument noise. The model prediction results (next section) will demonstrate how the FSS approach was able to perform quantification and differentiation of each component in the mixtures of AHA and GRW. Model predictions were generated using the validation data set (matrix X_v).

5.3.2 Model Predictions within Calibration Range

Model predictions of AHA concentrations in four different GRW concentrations obtained by dilution are compared with measured DOC values (Figure 5-2). Figure 5-3 contains a comparison of the prediction of GRW concentrations corresponding to the different dilutions examined. The points identified by dashed circles (Figure 5-2 and Figure 5-3) are the predictions obtained outside the calibration range of AHA. All predictions, that correspond to DOC values within the calibration range (0.5 - 5.0 mg-DOC/L), show very good agreement (Figure 5-2a-d). The maximum prediction error is in the order of 0.5 mg-DOC/L which corresponds to approximately 10% of the total variation of the AHA concentration. Model predictions are also quite accurate for the GRW equivalent DOC that fall within the calibration range (Figure 5-3). The maximum prediction error is in the order of 0.2 mg-

DOC/L which corresponds to approximately 5% of the total variation of the concentration of GRW. These maximum errors are only occurring for a small percentage of the cases tested. Prediction errors of AHA and GRW concentrations in many instances do not exceed 5% and 3%, respectively of the total variation in concentration values within the calibration range.

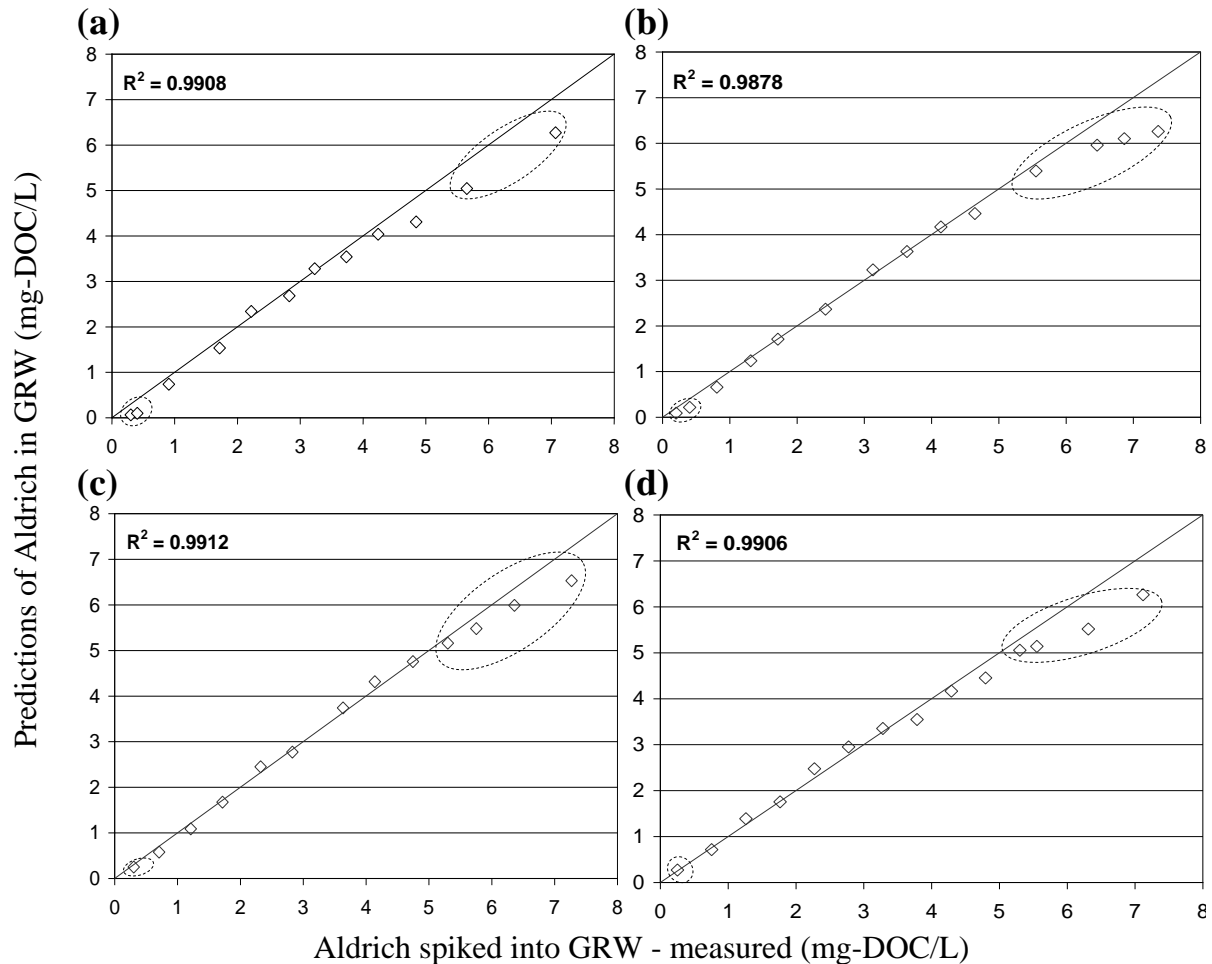


Figure 5-2 Predictions of the quantity of AHA mixed into GRW, diluted up to (a) 10%, (b) 25%, (c) 50% and (d) 75% of its original strength. Predictions outside the calibration range are circled. The straight line through the points, represents a perfect match between the measured and predicted values.

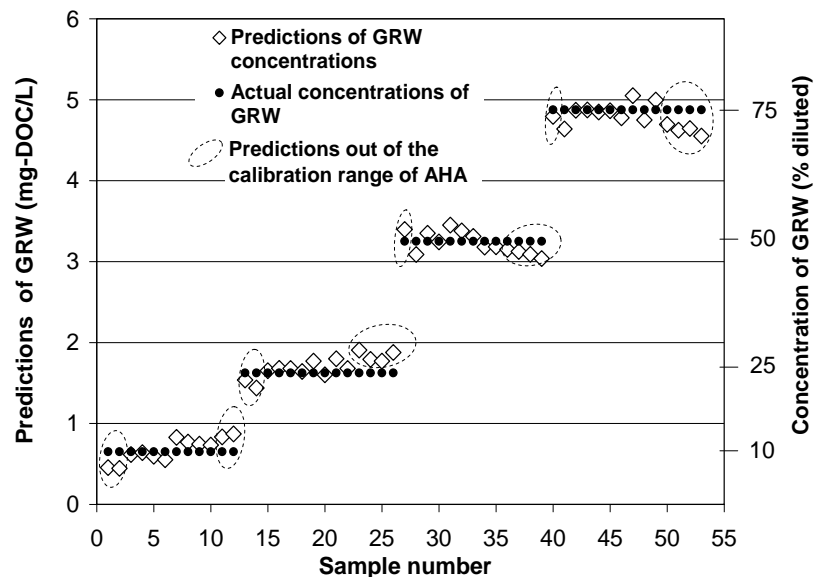


Figure 5-3 Predictions of the quantity of GRW in different mixtures of GRW and AHA. Predictions outside the calibration range of AHA are circled. The secondary vertical axis shows the comparable dilution levels (% of original strength) of GRW to the measured DOC values on the primary vertical axis.

5.3.3 Model Predictions outside Calibration Range

Although the predictions are very accurate for points within the calibration range, the accuracy degrades for conditions outside this range. The maximum prediction error, in this case, is however under 22% and 8% of the total variation in concentration of AHA and GRW, respectively (Figure 5-2a-d and Figure 5-3). The AHA concentrations, that exceeded the upper limit of the calibration range, were under predicted for all dilution levels of GRW. This reflects the fact that there is a non-linear correlation present in the fluorescence data, possibly due to inner filter effects, at higher AHA concentrations as manifested in terms of under predictions beyond the upper limit of the calibration range. A further discussion on this non-linear behaviour is presented in Supplementary Information (SI) Section 5.5.1. The PLS algorithm used in this study can only generate models that have linear combinations of LVs.

Therefore, the calibration model's inability to handle non-linear behaviour seen here is generally acceptable.

In spite of the observed shortcomings for concentrations outside of the calibration range, prediction results demonstrate that the calibration model, generated by PLS regression, was able to differentiate and predict the individual AHA and GRW concentrations with good accuracy inside the calibration range. The model also showed good predictive abilities considering the overlapping of the fluorescence EEM of AHA and GRW examined in this study (Figure 5-1a-c) which generally makes it rather difficult to predict their individual contribution when these species are present in a mixture. The performance of the calibration model, using DOC for two different sources of NOM, illustrates the potential of this FSS approach for fast and reliable quantification of NOM in water. The limitation of this model in predicting concentrations outside the calibration range may be improved by adopting a non-linear quadratic PLS algorithm (Wold *et al.*, 1989). A non-linear PLS would require a larger number of calibration coefficients resulting in increased sensitivity to noise. Since the calibration range selected in this study covers typical DOC levels for natural surface waters, no effort was made to improve model predictions outside the calibration range.

5.3.4 Physical Relevance of Latent Variables

The calibration model developed in this study is a linear combination of three LVs. The predictor variables (spectral data) in the predictor data set (X) are linearly combined to form each LV by assigning appropriate weight values (weight loadings) to all predictor variables. This process therefore establishes a corresponding weight matrix (W) for each LV. Hence, by

examining W matrices, one could understand which original predictor variables in X space, i.e. which combinations of excitation and emission wavelengths would be most important within the LVs (Wold *et al.*, 2001). Since high weights are assigned to the variables in X space that are highly correlated with Y variables (response data set), the weight loading matrices can be used to understand which variables in the X space, i.e. intensities of the fluorescence EEMs at specific combinations of excitation and emission wavelengths, explain most of the variability in Y space, i.e. the DOC measurements.

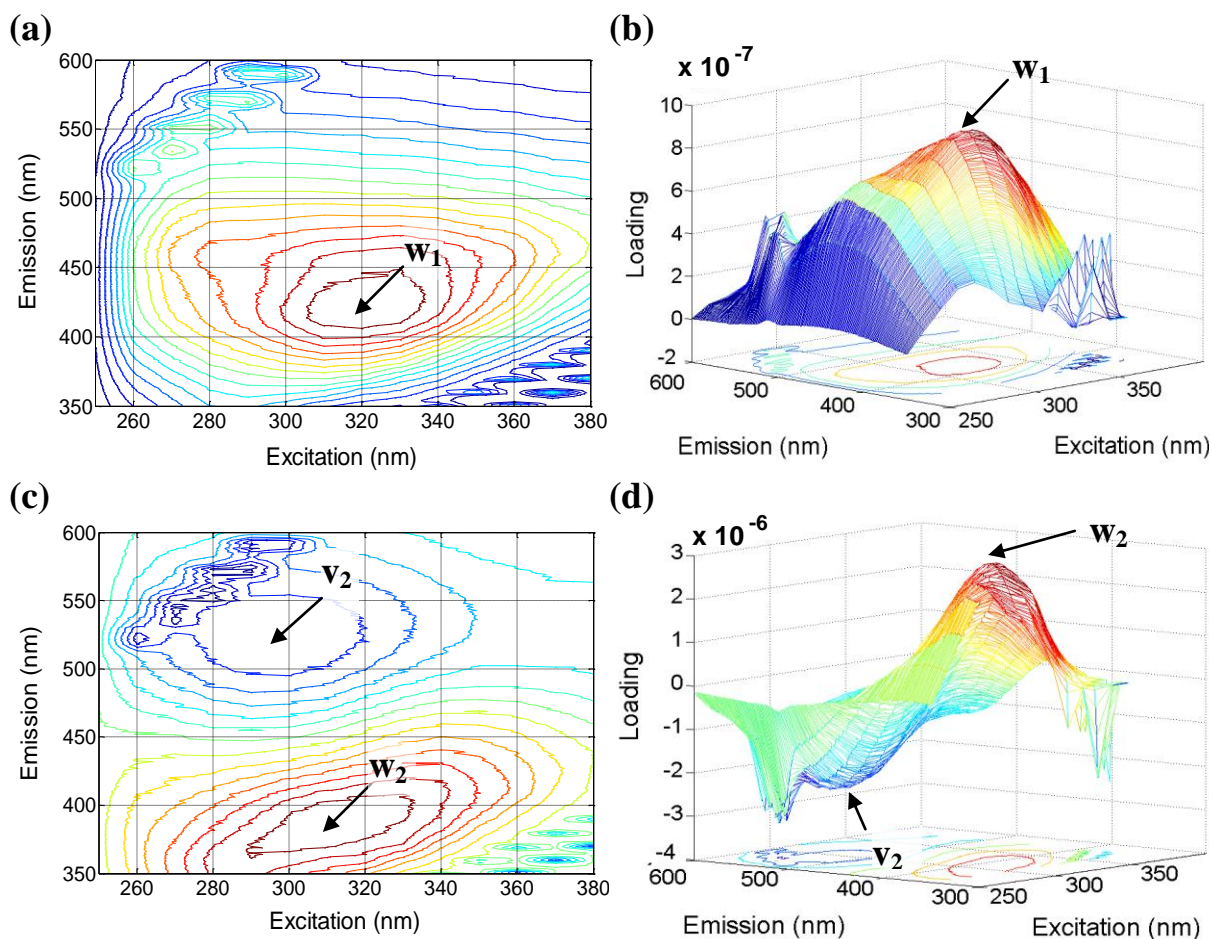


Figure 5-4 Weight loading values plotted at their corresponding excitation/emission wavelengths; (a) weight loading matrix on the first latent variable (LV1), (b) 3D representation of weight loadings on LV1, (c) weight loading matrix on the second latent variable (LV2) and (d) 3D representation of weight loadings on LV2.

When the weight loading values of the first latent variable (LV1) were plotted (overlaid) on the corresponding excitation and emission wavelengths of EEM, a weight loading peak (w_1) at Ex/Em \sim 320 nm/420 nm was observed (Figure 5-4a & b). An identical weight loading peak for LV1 was obtained when the same spectral data were correlated with the total (AHA and GRW) concentrations of the mixtures using PLS regression (results not shown). Therefore it can be concluded that the peak (w_1) and consequently LV1 is dominant in explaining the total DOC concentration of the sample mixture. The weight loading values of the second latent variable (LV2), plotted following the same procedure, shows a peak (w_2) and a valley (v_2) at Ex/Em \sim 310 nm /380 nm and Ex/Em \sim 295 nm /525 nm respectively (Figure 5-4c & d). The peak (w_2) and the valley (v_2) are situated in the corresponding neighbourhoods of the main peaks observed for the individual spectra of GRW and AHA. In addition, when the calibration model had only one LV (LV1), no differentiation of AHA and GRW was possible. As a result, it can be concluded that LV2 is responsible for differentiating the concentrations of AHA and GRW from the concentration of the total mixture by assigning large positive weights for the GRW region and large negative weights to the AHA region of the fluorescence EEM. A similar distribution of weights was seen in the weight loading matrix of the third latent variable (LV3) with a peak (w_3) and a valley (v_3) at Ex/Em \sim 280 nm /350 nm and Ex/Em \sim 310 nm /525 nm, respectively (results not shown). The inclusion of LV3 was necessary to improve the model accuracy as deduced from the improvements observed in the root mean square errors of the model predictions with three latent variables against two latent variables. Both root mean square error of calibration (RMSEC) and cross validation (RMSECV) decreased as the number of LVs used in the model was increased. The number of LVs was limited to three as the inclusion of the fourth

LV caused the RMSECV value in the prediction of GRW concentrations to increase (results not shown).

5.3.5 Significance of the FSS Approach Compared to Other Fluorescence-based Methods

A similar PLS-based model developed using the FRI approach (Chen *et al.*, 2003) to quantify AHA and GRW concentrations resulted in model predictions that were significantly less accurate compared to the FSS approach (SI Section 5.5.2). In addition, the calibration model developed using the FSS approach was found to be less sensitive to measurement noise than a model obtained with the FRI method (SI Section 5.5.2).

The fluorescence peak-picking method was unable to de-convolute the non-linearly correlated EEM intensities of AHA and GRW and was therefore not suitable for quantification and differentiation of AHA and GRW concentrations as demonstrated in SI Section 5.5.3.

5.4 Conclusions

This study demonstrates how a calibration model developed using the FSS approach and PLS regression was able to quantify and differentiate HA- and FA-like NOM present in a complex natural water matrix. This was demonstrated by using AHA and GRW to represent HA- and FA-like NOM, respectively in this study. In particular, this approach was suitable for quantifying AHA and GRW concentrations in different mixtures with good accuracy for

conditions within the limits of the calibration range. The lower accuracy of the model predictions of AHA and GRW concentrations outside the upper limit of the calibration range was attributed to the non-linear correlation present in the fluorescence data. Compared to other available NOM characterizing methods, the FSS approach shows better accuracy and is less sensitive to noise. This proposed approach provides a promising methodology for the quantification of NOM in natural water.

5.5 Supplementary Information

5.5.1 The Non-linear Behaviour of the Predictions at High AHA Concentrations

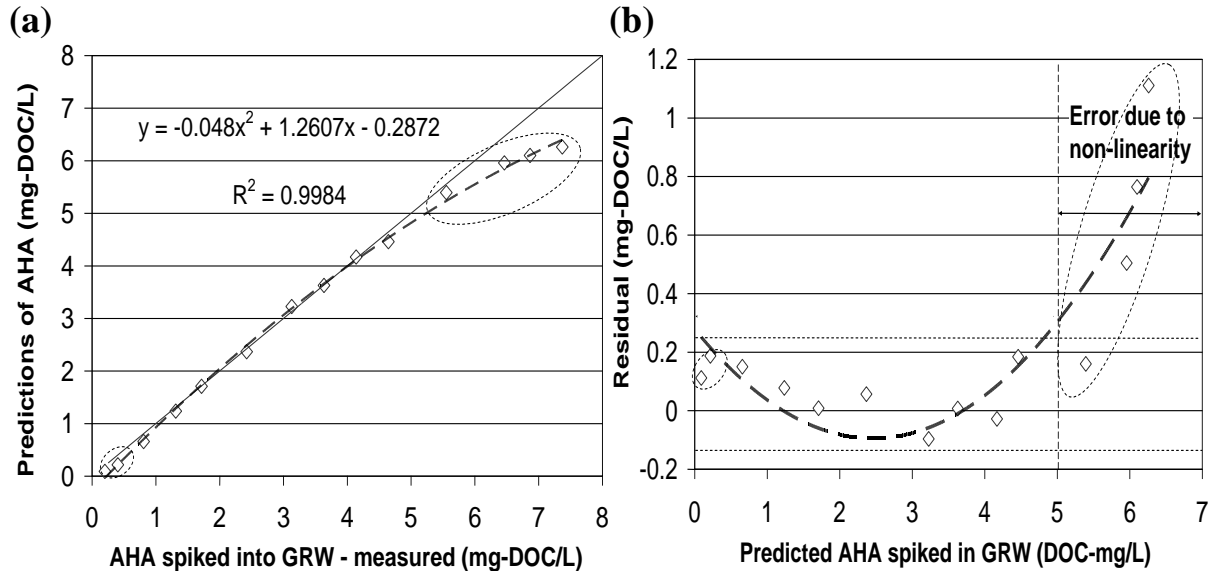


Figure S - 5-1 An example of non-linear behaviour manifested in the predictions of AHA: (a) prediction of AHA concentration represented as function of calculated AHA concentration and (b) residual plot of model predictions. Predictions outside the calibration range are circled.

The non-linear behaviour present at higher AHA concentrations is further evident when Figure 5-2 in the manuscripts is closely examined. In particular, it could be shown that the predicted AHA concentration can be represented as a second order function of estimated AHA concentration for all GRW dilutions considered. Figure S - 5-1a demonstrates this phenomenon with the model predictions for AHA mixed into GRW, diluted up to 25% of its original strength. In addition, the residual plot (Figure S - 5-1b) which demonstrates the prediction error against the model predictions of AHA mixed into GRW, diluted up to 25% of its original strength, shows how the prediction error is distributed randomly below the upper limit of the calibration range between 0.2 and -0.2 mg-DOC/L and how it generally increases above the upper limit of the calibration range. Comparatively poor predictions in

this range are attributed to the absence of a quadratic term in the calibration model as indicated by the parabolic trend line drawn on the residual plot (Figure S - 5-1b). Similar observations were obtained for all other dilution levels of GRW (results are not shown).

5.5.2 Quantification of AHA and GRW using a Calibration Model Based on FRI Approach

The physical relevance of LV2 is related to the differentiation of AHA and GRW concentrations from the total mixture concentrations. To demonstrate this phenomenon in a more practical perspective, the FRI approach (Chen *et al.*, 2003) was used to develop a similar calibration model as presented in this study using PLS and the full EEM spectrum. Another objective of using the FRI approach was to compare the accuracy of the predictions obtained using the fluorescence-based soft sensor approach with the FRI approach.

In the FRI approach, the EEMs of all AHA and GRW mixture samples were divided into two regions (AHA-region and GRW-region) based on the established peak locations of AHA and GRW as shown in Figure S - 5-2a. The exact positions of this division were based on the line that divides the peak and valley regions in the weight loading matrix of the LV2 as demonstrated in Figure S - 5-2b. The quantification of AHA and GRW was then based on the volume beneath the EEMs of AHA and GRW mixture samples using Equation S – 5.1. Using $I(\lambda_{\text{ex}}, \lambda_{\text{em}})$, the fluorescence intensity at each excitation-emission wavelength pair, the volume (V_i) beneath region “i” of the EEM can be calculated with

$$\sum_{ex} \sum_{em} I(\lambda_{ex}, \lambda_{em}) \Delta \lambda_{ex} \Delta \lambda_{em} \quad (S - 5.1)$$

where $\Delta \lambda_{ex}$ is the excitation wavelength interval (taken as 10 nm) and $\Delta \lambda_{em}$ is the emission wavelength interval (taken as 1 nm). The volumes calculated under the AHA-region, the GRW-region and the total EEM of each sample were correlated with the corresponding DOC content of GRW and AHA to develop a calibration model using PLS regression.

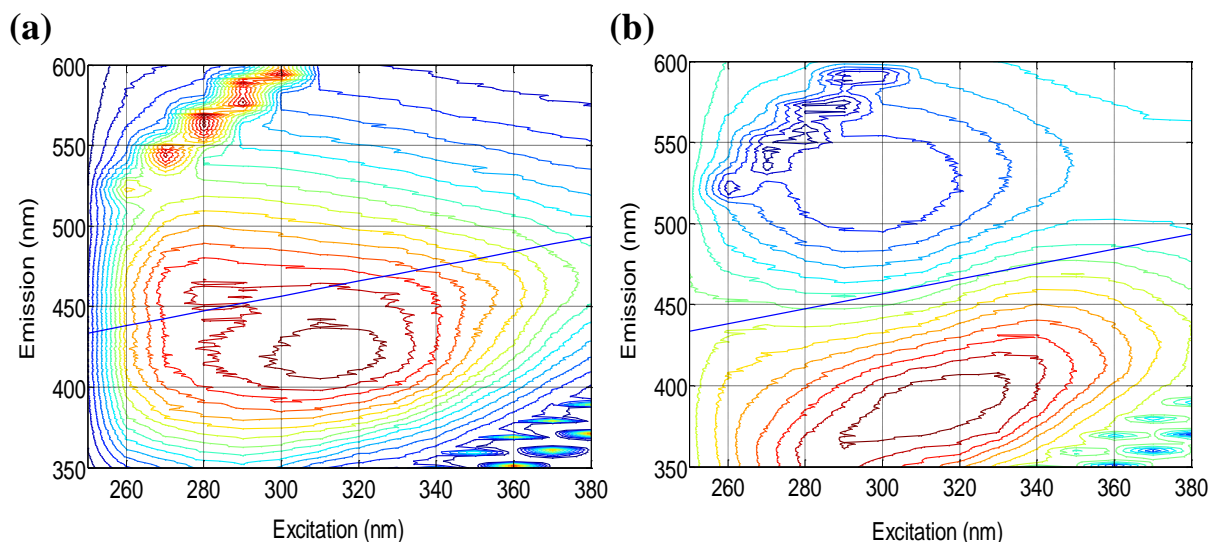


Figure S - 5-2 Two regions of the fluorescence EEM used in the FRI analysis. (a) EEM of GRW and (b) weight loading matrix of LV2.

The model developed using the FRI approach was able to quantify AHA with reasonable accuracy while the FRI model predictions of GRW concentrations were comparatively less accurate in most cases (Table S - 5-1). Nevertheless, the ability of the FRI-based model to differentiate AHA and GRW demonstrates that the division which was based on the peak and valley present in the weight loading matrix of LV2 was reasonable and further strengthens the argument regarding the physical relevance of LV2. However, FRI-based model predictions are significantly less accurate compared to the fluorescence-based soft sensor

approach (Table S - 5-1). This can be explained by the fact that the latter approach analyses every data point (smallest sub-region) of the EEM capturing most of the variability present in the fluorescence spectra while the FRI approach only lumps the information of large regions into one independent variable thus losing some of the information contained in the data. Therefore the FRI approach is only able to capture the variability present in the data to a limited extent.

Table S - 5-1 Comparison of model prediction error in terms of the total variation in DOC concentration for GRW and AHA between fluorescence-based soft sensor and FRI approaches.

	Prediction error (%)			
	Fluorescence-based soft sensor		FRI	
	AHA	GRW	AHA	GRW
Inside calibration range				
Maximum error	10	5	24	22
Error in most cases	< 5	< 3	< 12	< 18
Outside calibration range				
Maximum error	22	8	36	20

In addition, the sensitivity of both methods to measurement noise was examined by introducing random $\pm 5\%$ artificial measurement noise for fluorescence EEM intensities of randomly selected samples covering the entire validation data set (X_v). When these fluorescence data with artificial measurement noise were used for predictions, the FRI-based method performed poorly compared to the fluorescence-based soft sensor method within the calibration range. In particular, with the fluorescence-based soft sensor method, the maximum prediction error of AHA and GRW concentrations did not change significantly from the values given in Table S - 5-1. However, with the FRI-based method, the prediction

accuracy of AHA and GRW concentrations was greatly affected (results are not shown). Therefore it can be concluded that the fluorescence-based soft sensor method is less sensitive to measurement noise compared to the FRI-based method.

5.5.3 Significance of Fluorescence-based Soft Sensor Approach Compared to Analyzing Only Major Ex/Em Peak Positions (Peak-Picking Approach)

When only the Ex/Em peak intensities of AHA and GRW were taken into account, the intensity readings were not linearly correlated with the concentrations (Figure S - 5-3a and b). Hence, using major Ex/Em peak positions, it is impossible to develop a calibration model that could quantify the concentrations of the individual substances independently from a mixture of GRW and AHA. PLS regression is able to solve this problem by de-convoluting these non-linearly correlated EEM intensities captured using the fluorescence-based soft sensor approach.

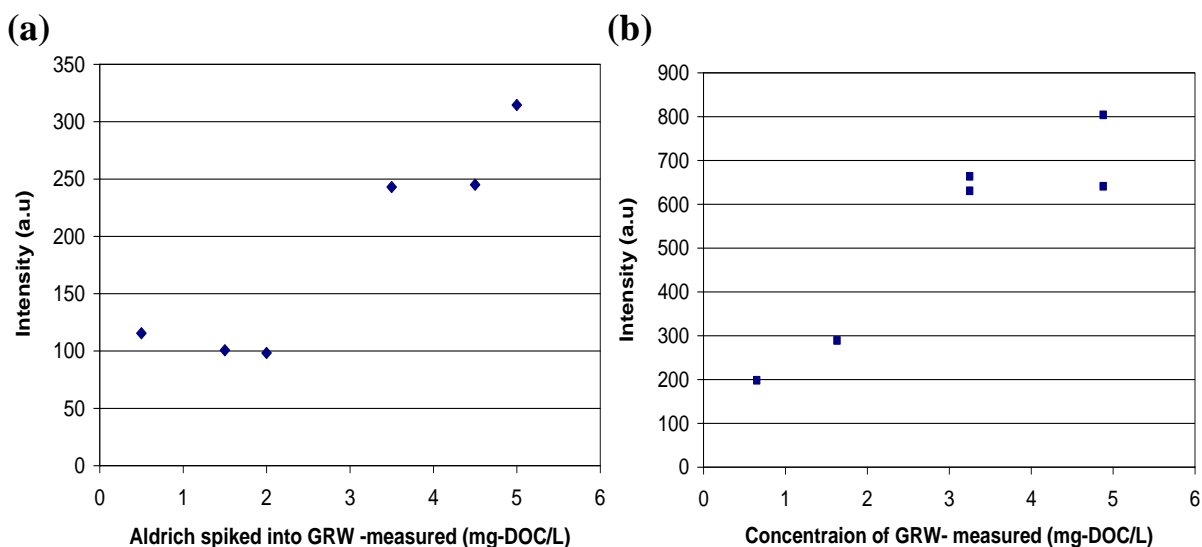


Figure S - 5-3 Comparison of main peak intensities of (a) AHA (Ex/Em: 270nm /460 nm) and (b) GRW (Ex/Em: 320nm /415 nm) at different concentration levels for randomly selected samples in the calibration set presented in this chapter.

CHAPTER 6

Fluorescence-based Rapid Identification of Different Natural Organic Matter Fractions for Drinking Water Treatment Applications

Overview

Rapid and accurate monitoring of natural organic matter (NOM) in drinking water treatment and distribution systems is essential for maintaining efficient water treatment processes and protection of public health. The individual and combined contributions of different NOM fractions, such as protein-, humic acid-, fulvic acid- and polysaccharide-like substances, present in natural water contribute to membrane fouling, disinfection by-products (DBP) formation and undesirable biological growth in drinking water treatment processes and distribution systems. Most of the NOM characterization methods that are available are not suitable for differentiating all these NOM fractions and are not amenable to rapid analyses. In this research, a fluorescence excitation-emission matrix (EEM)-based calibration model was developed as a rapid NOM identification method. Partial least squares regression (PLS) was used to develop this calibration model by correlating the fluorescence intensities of water samples that contained surrogate NOM fractions with their corresponding dissolved organic carbon concentrations. Natural river water, Aldrich humic acid, α -lactalbumin and alginic acid were used as surrogate NOM fractions to represent fulvic-, humic-, protein- and polysaccharide-like NOM fractions, respectively. The proposed approach was found to be

suitable for accurately identifying changes of different NOM fractions for the water samples used in this study. In addition, the feasibility of this approach for identifying the extent of interaction between polysaccharide- and protein-like NOM fractions was also demonstrated. Hence, this proposed methodology could have applications in identifying/monitoring the extent of the relative changes that occur in different NOM fractions in water and the interactions between polysaccharide- and protein-like NOM in water treatment processes and distribution systems.

Keywords: drinking water treatment, fluorescence spectroscopy, natural organic matter, membrane foulants, partial least squares regression.

6.1 Introduction

Rapid monitoring of natural organic matter (NOM) for various drinking water applications is becoming increasingly important to ensure the reliability of water treatment processes and distribution system performance in the context of maintaining water treatment process efficiency and protection of public health. For example, NOM in water causes membrane fouling in membrane-based drinking water treatment due to the accumulation of NOM on the surface and/or in the pores of the membranes. As a result, this increases the operational costs due to increased pressure requirements and the cost associated with increased energy, backwashing and chemical cleaning requirements. NOM also promotes undesirable biological growth in drinking water distribution systems (Christy *et al.*, 2000) and increases the potential for the formation of disinfection by-products (DBP) during disinfection/oxidation of the water (Marhaba *et al.*, 1998). These negative effects may occur

as a result of the individual and combined contributions of major NOM fractions such as protein-like, humic acid-like, fulvic acid-like and polysaccharide-like substances. Humic acid- and fulvic acid-like NOM fractions have a greater potential for forming DBP compared to protein- and polysaccharide-like NOM fractions (Marahaba *et al.*, 2000). Also, the interplay between the major NOM fractions could lead to different membrane fouling behaviour during the production of drinking water (Jermann *et al.*, 2007).

Unpredictable rainfall and snow melting patterns that originate due to short-term changes in weather conditions can often be the source for rapid changes in NOM composition and concentration levels. Rapid and accurate monitoring tools that could be used to identify the changes in different NOM fractions are required for the development and the implementation of strategies to mitigate the negative effects of NOM for drinking water treatment and distribution systems.

The methods available for rapid monitoring of water treatment (drinking and wastewater treatment) systems such as conductivity, total organic carbon (TOC) content and, transmembrane pressure for membrane filtration systems (Henderson *et al.*, 2009) do not provide detailed information related to the changes in the characteristics of NOM in water. Most available NOM characterization methods (Huber *et al.*, 1992; Her *et al.*, 2003; Gray *et al.*, 2007) are neither suitable for differentiating the major NOM fractions nor can they be used for rapid analysis. Furthermore, these methods are generally expensive and lack the sensitivity to provide direct and accurate analyses of very low NOM concentration levels as found in nanofiltration (NF) permeate (Peiris *et al.*, 2009). In this context, fluorescence

spectroscopy is gaining significant attention for the characterization of NOM and shows good potential for on-line or rapid off-line monitoring, as minimal sample pre-treatment and preparation is needed and high instrumental sensitivity is available (Coble *et al.*, 1990; Peiris *et al.*, 2008).

Various fluorescence spectroscopic techniques have been used to study NOM characteristics in water (Coble *et al.*, 1990; Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003; Sierra *et al.*, 2005; Hudson *et al.*, 2007; Henderson *et al.*, 2009). Fluorescence excitation-emission matrices (EEM) of natural water contain specific fluorescence information that correlate with humic-, fulvic- and protein-like species' concentrations in terms of fluorescence intensities at different excitation wavelength (Ex) and emission wavelength (Em) combinations. The light scattering regions present in the fluorescence EEMs can also provide information related to the particulate/colloidal matter present in water (Wyatt, 1993; Stramski *et al.*, 2005; Peiris *et al.*, 2010). Unlike the fluorescence single scan approach (i.e. collecting data for only the maximum fluorescence peak), fluorescence EEM approach has been shown to be capable of capturing the subtle changes that occur in the NOM content of water over a wide range of dissolved organic carbon (DOC) concentrations that arise due to seasonal effects or other changes, without pre-dilution or pre-concentration steps (Peiris *et al.*, 2009).

Most of the reported fluorescence EEM-based techniques for NOM characterization have considered only major fluorescence peak positions (peak-picking method) out of the entire fluorescence spectrum that contain thousands of wavelength-dependent fluorescence intensity data points (Chen *et al.*, 2003). These techniques have been found lacking in their ability to capture the heterogeneity of the different NOM fractions in water and therefore, do

not provide accurate characterization of specific NOM fractions. Also, the use of the full fluorescence EEM as opposed to using only the main peak positions provides better sensitivity in terms of identifying the changes that occur in different NOM fractions and particulate/colloidal matter in water (Peiris *et al.*, 2010). Some reported characterization techniques, such as fluorescence regional integration (FRI) (Chen *et al.*, 2003) and the PARAFAC approach (Stedmon *et al.*, 2003), do take into account all the data points in the full fluorescence EEM and are therefore better suited for NOM characterization in water. However, these methods do not characterize all the above mentioned major NOM fractions especially the polysaccharide-like NOM or the interactions between different NOM fractions.

In this study, a fluorescence EEM-based approach is proposed that is suitable for differentiating the major NOM fractions in water and for estimating the extent of relative changes in these NOM fractions that occur for different water treatment processes. The advantage of the approach taken here is that it can be used to acquire this information quickly and to characterize polysaccharide-like NOM and its interactions with protein-like NOM. The proposed approach could serve as a rapid monitoring tool for estimating the relative changes in NOM fractions present in water, which would be very helpful for researchers, plant operators and engineers who are interested in studying the effects of NOM in drinking water treatment applications and distribution systems, and for developing monitoring and control strategies to better manage these systems.

This approach is based on earlier work for acquiring reproducible fluorescence spectra for NOM at very low concentrations (Peiris *et al.*, 2009). In the current study, a calibration

model was developed by correlating fluorescence spectroscopic data with independent DOC measurements for four different surrogates representing the major fractions found in NOM. The interactions between the major NOM fractions are also discussed. Partial least squares regression (PLS), a multivariate statistical technique, was used for developing the calibration model which correlates the intensities captured within the full fluorescence EEMs of water samples containing different concentration levels of NOM fractions with their corresponding DOC concentrations. The use of PLS is justified by its ability to handle the high collinearity present among fluorescence intensity measurements captured across the EEM (Elshereef *et al.*, 2006).

6.2 Methods

6.2.1 Experimental Approach and Rationale

Surrogates were selected that represent the four major NOM fractions found in natural water. Grand River water (GRW) (Southwestern Ontario, Canada; collected on June 22, 2008) was selected to represent fulvic-like NOM in water as river water sources have been reported to contain predominantly fulvic acid-like NOM (Huck, 1999; Sierra *et al.*, 2005); Aldrich humic acid (AHA), obtained from Sigma Aldrich Inc., (St. Louis, MO; batch #: 04013JD, humic acid, sodium salt), was chosen to represent humic acid-like NOM in natural water. Even though AHA is largely a soil-based substance (Malcolm and MacCarthy, 1986), it is reasonable to assume that humic acid-like NOM present in natural water takes its origin from the soil closer to the water sources. To represent protein-like and polysaccharide-like NOM fractions in water, α -lactalbumin (AL) and alginic acid (AA), respectively were selected. AL,

a major whey protein, was donated by Davisco Foods International, Inc., (LeSueur, MN; Lot #: JE 007-3-921). AA was obtained from Sigma Aldrich Inc., (St. Louis, MO; Lot #: 100K0219, alginic acid, sodium salt). GRW, AHA and AL demonstrated similar fluorescence spectral properties to that of fulvic-like, humic-like and protein-like NOM, respectively that are present in natural water (results discussed in detail later). AA does not fluoresce, and hence does not have a specific fluorescence signature. AA has often been used as a surrogate polysaccharide-like NOM in water research (Kweon *et al.*, 2005; Jeremann *et al.*, 2007). Its selection could be further justified by examining its liquid chromatography - organic carbon detection (LC-OCD) chromatograms (results discussed later). Different amounts of AHA, AL and AA were then spiked at different concentration levels of GRW (made by dilution) to prepare calibration and validation sample sets as explained below.

The rationale in using this approach is that when the concentrations of surrogate NOM fractions change within each sample, spectral properties captured in the fluorescence EEM also change in relation to the changes in the individual NOM concentrations. Since surrogate NOM fractions used in this study demonstrate similar properties to the NOM fractions found in natural water, it could be expected that the calibration model that relates these spectral changes to the changes in the DOC content of surrogate NOM fractions could be used as a model to identify and estimate the changes in the NOM fractions that occur in natural water.

6.2.2 Sample Preparation.

Stock solutions of AHA, AL and AA, each about 1 g/L, were made by dissolving 1 g each of AHA, AL and AA, in Milli-Q (Millipore) water (resistivity = 18.2 M Ω cm). For sample

preparation, these stock solutions were subsequently diluted in Milli-Q water to prepare 20 mg/L AHA, AL and AA solutions. All the samples withdrawn from the stock solutions to prepare these diluted AHA, AL and AA solutions were filtered using 0.45 μ m Nylon Acrodisc type syringe filters to remove particulate matter. For the same reason, GRW was also filtered through 0.45 μ m GN-6 Metrical® type 47 mm disc filters.

6.2.3 DOC and LC-OD analyses

The DOC contents (mg/L) of GRW, AHA, AL and AA solutions were recorded using an OI-Analytical TOC analyzer (Model 1010, College Station, TX) using a wet-oxidation method as described in Standard Methods (2005) 5310D. LC-OCD analyses were performed on a DOC-Labor, Dr Huber (Karlsruhe/Germany) system (Huber and Frimmel, 1992).

6.2.4 Fluorescence Analysis

The fluorescence EEMs of each sample was recorded using a Varian Cary Eclipse Fluorescence Spectrofluorometer (Palo Alto, CA) collecting 301 individual emission intensities (Em: 300 - 600 nm) at sequential 10 nm increments of excitation wavelengths between 250 and 380 nm. UV-grade polymethylmethacrylate (PMMA) cuvettes with four optical windows (path length: 10 mm) were used in the analyses. Instrument parameters: photomultiplier tube (PMT) voltage = 800 V, scan rate = 600 nm/min and excitation/emission slit width = 10 nm each were maintained during the fluorescence signal acquisition. These parameter settings have been previously identified as optimal for obtaining reproducible fluorescence signals, especially for low NOM concentrations (Peiris *et al.*, 2009). Corrections for inner filter effects were not applied as inner filtering is unlikely to

occur at the low concentration levels examined in this study (Hudson *et al.*, 2008). To eliminate water Raman scattering and to reduce other background noise, fluorescence spectra for Milli-Q water, obtained under the same conditions, were subtracted from all spectra. During the course of the fluorescence analyses, the Raman scattering peak intensity recorded for Milli-Q water at Ex/Em ~ 348 nm/396 nm was less than 1%, confirming that there was no significant fluctuations in the performance of the spectrofluorometer lamp or other hardware. The samples were maintained at room temperature (~ 25 °C) during the analyses. The pH of all the samples was adjusted to ~ 8.2 to match the pH of GRW using high purity 0.1 M HCl acid (99.999%) and 0.1 M NaOH (99.998 %) before fluorescence analysis.

6.2.5 Calibration Set

Four sets of sample mixtures, with 15 samples in each set, were made with each set containing different amounts of AHA, AL and AA and a fixed amount of GRW. Different concentrations of GRW for a given set were obtained by dilution of GRW (10%, 25%, 50% and 75% of the original strength or 0.65, 1.63, 3.25 and 4.88 mg-DOC/L, respectively). The concentration of AHA, AL and AA of each sample in all four sets were varied to achieve the concentration levels illustrated in the experimental cube design (ECD) (Figure 6-1). The DOC content of GRW, AHA, AL and AA in each sample was calculated from the DOC content (mg-DOC/L) of the original GRW, AHA, AL and AA solutions used for sample preparations. To examine the reproducibility of the sample preparation method, sample triplicates were made from each set that had AHA, AL and AA concentration combinations corresponding to the centre point of the ECD (Figure 6-1).

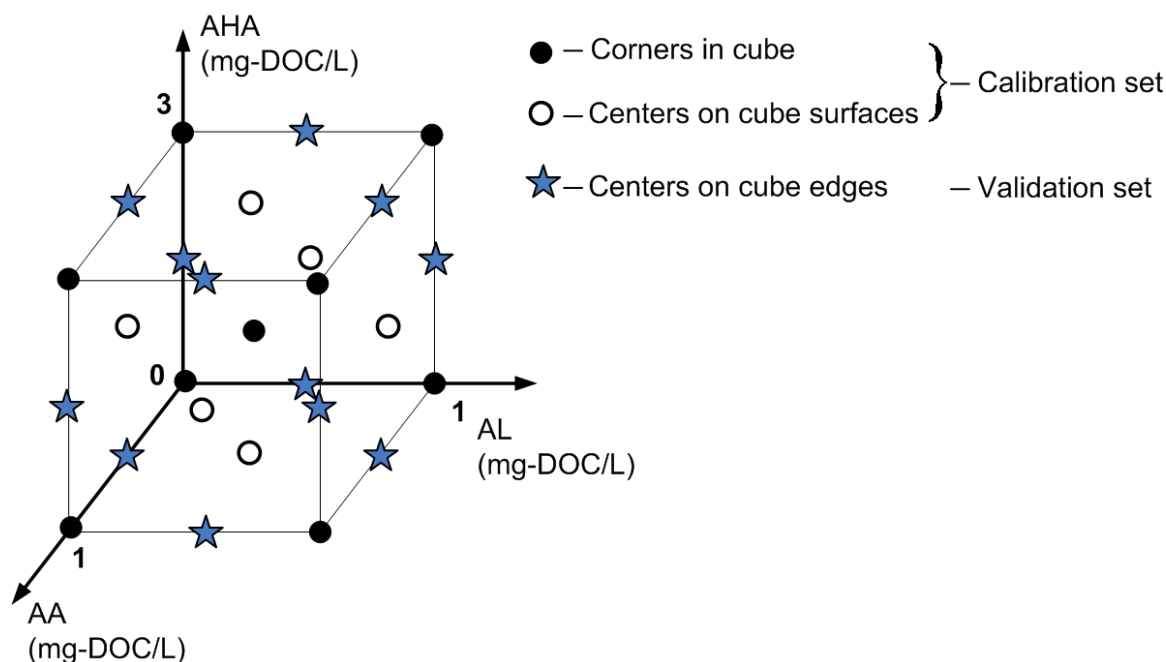


Figure 6-1 Experimental cube design (ECD) used in the sample preparation for calibration and validation sample sets. Three factors: Aldrich humic acid (AHA), alginic acid (AA) and (α -lactalbumin) AL are shown here. The fourth factor: Grand River water (GRW), not shown, has concentration levels of 0.65, 1.63, 3.25 and 4.88 mg-DOC/L.

6.2.6 Validation Set

In addition to the calibration set, four sets of samples (12 samples in each set) were prepared, with each set having one of the four dilution levels of GRW that were used in the calibration samples. AHA, AL and AA concentration levels of each sample in each set were determined by the center points at the edges of the ECD (Figure 6-1). The total DOC (i.e. DOC of AHA, AA, AL and GRW) in both calibration and validation sample sets were varied from ~ 0.65 – 10 mg-DOC/L to cover the typical DOC levels found in natural surface waters (Thurman, 1985). The individual concentration ranges of AHA, AL and AA were selected to cover typical concentration levels of humic acid-like, protein-like and polysaccharide-like NOM found in natural waters (Thurman, 1985; Hallé *et al.*, 2009).

6.2.7 Analysis of Variance (ANOVA) of Fluorescence EEMs

To understand how the fluorescence EEM intensities are affected due to the changes in different NOM fractions in water, analysis of variance (ANOVA) was performed on the fluorescence EEM intensity data. For this purpose, fluorescence EEMs captured within a 2^4 factorial design which contained two concentration levels each of GRW, AHA, AL and AA were used. For each of the two concentration levels of GRW (1.63 and 3.25 mg-DOC/L), concentrations of AHA, AL and AA were varied according to the levels illustrated by the closed circles in the ECD in Figure 6-1. The fluorescence EEM intensity data of the samples that correspond to the center point of the ECD (Figure 6-1) were used to calculate experimental error that was needed in the ANOVA analysis.

6.2.8 Model Development and Validation

The fluorescence EEM of each sample contained intensity readings at 4214 excitation and emission coordinate points (i.e. 4214 spectral variables). All of these intensity readings of each EEM were rearranged to generate data rows of intensity values as illustrated in the supplementary information (SI) section (Figure S - 6-1). This procedure generated a 60×4214 data matrix (X_c) from the calibration sample set and a 48×4214 data matrix (X_v) from the validation sample set. Each row of these data matrices corresponded to a sample and the intensity values of the corresponding EEM were arranged over 4214 columns. The DOC values of AHA, AL, AA and GRW for each sample of the calibration sample set was also arranged into a 60×4 data matrix (Y_c) with each row containing the DOC values of AHA, AL, AA and GRW.

The X_c and Y_c data matrices were then used as the predictor data set and the response data set, respectively to develop a calibration model using PLS (SI Figure S - 6-1). PLS analysis is a well-known technique for developing calibration models which can be used to correlate between multiple explanatory variables such as fluorescence spectral variables in X_c and multiple response variables such as NOM concentrations measurements in Y_c . PLS is also able to handle problems associated with the noise and collinearity present among these fluorescence spectral variables (Elshereef *et al.*, 2006). Essentially, PLS analysis extracts a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and mathematically represented by linear combinations of original predictor variables. These linear combinations, referred to as latent variables (LVs), are calculated to account for as much of the manifest covariance as possible between predictor data set (X_c) and the response data set (Y_c) (Qin *et al.*, 1992). Both X_c and Y_c data sets were mean-centered before PLS analysis. The optimum number of LVs and the goodness of predictions (Q^2) were determined by the leave-one-out cross-validation algorithm (Eriksson *et al.*, 2001; Wold *et al.*, 2001). The validation data set (X_v) was subsequently used on the developed calibration model to validate the accuracy of model predictions. All computations were performed using PLS Toolbox 3.5 (Eigenvector Research, Inc., Manson, WA) within the MATLAB 7.3.0 computational environment (MathWorks, Natick, MA).

6.3 Results and Discussion

6.3.1 LC-OCD Characterization of Major Surrogate NOM Fractions

Comparison of Grand River water against the surrogate NOM fractions used in the development of the calibration model are shown in Figure 6-2. The LC-OCD chromatogram of GRW has two main peaks that correspond to humic substances (HS) and biopolymers at elution times of 50 and 40 minutes, respectively (Figure 6-2a). These were found to be similar to most LC-OCD chromatograms reported for natural water (Hallé *et al.*, 2009; Zheng *et al.*, 2009). Because HS have a lower molecular weight (MW) than the biopolymers, the elution time is longer (Figure 6-2a). HS consists of fulvic acid-like and humic acid-like NOM but the LC-OCD approach is not able to differentiate these two NOMs. The biopolymers found in natural water are comprised of polysaccharide and protein-like NOM, and as with HS, LC-OCD is also not suitable for differentiating these components (Figure 6-2a).

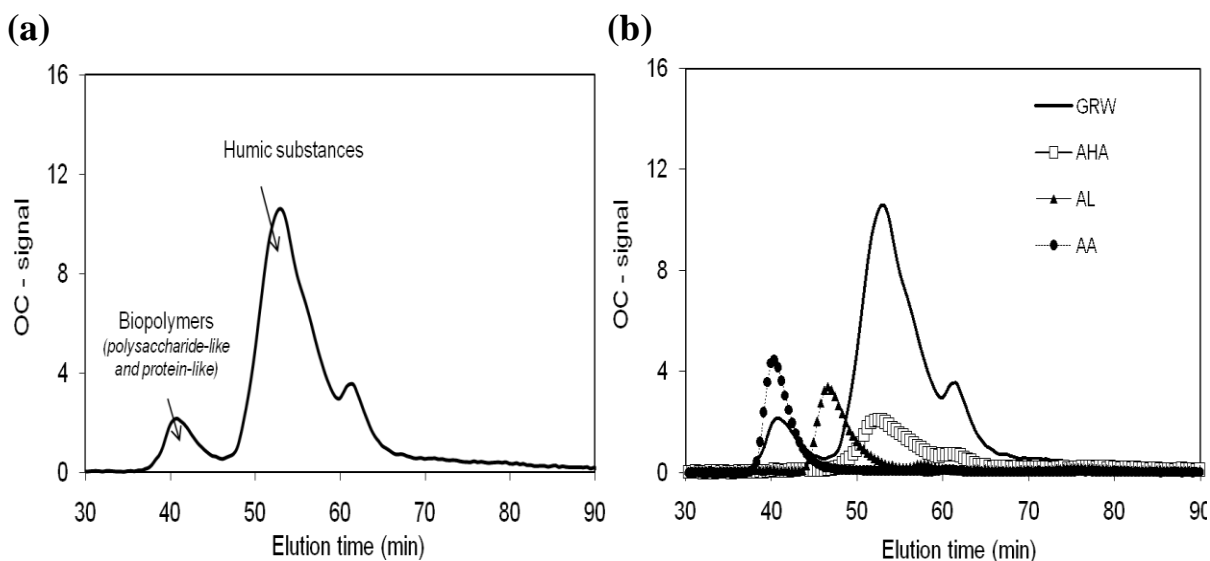


Figure 6-2 LC-OCD chromatograms of (a) Grand River water and (b) surrogate NOM fractions used in the development of the fluorescence-based calibration model.

AHA and AA are situated at the same locations of the elution profile as humic-acid like and polysaccharide-like NOM in GRW (Figure 6-2a). This corroborates that AHA and AA have a similar MW distribution to that of humic-acid like and polysaccharide-like NOM. The average MW of AL, on the other hand, appears to be a little lower than the protein-like NOM and closer to HS present in natural waters as demonstrated by the LC-OCD peak position of AL in Figure 6-2b. This is however expected as the MW distribution of protein-like matter present in natural waters can also include the upper limit of the MW distribution of HS (Jiang, 2007).

6.3.2 Fluorescence EEMs of Surrogate NOM

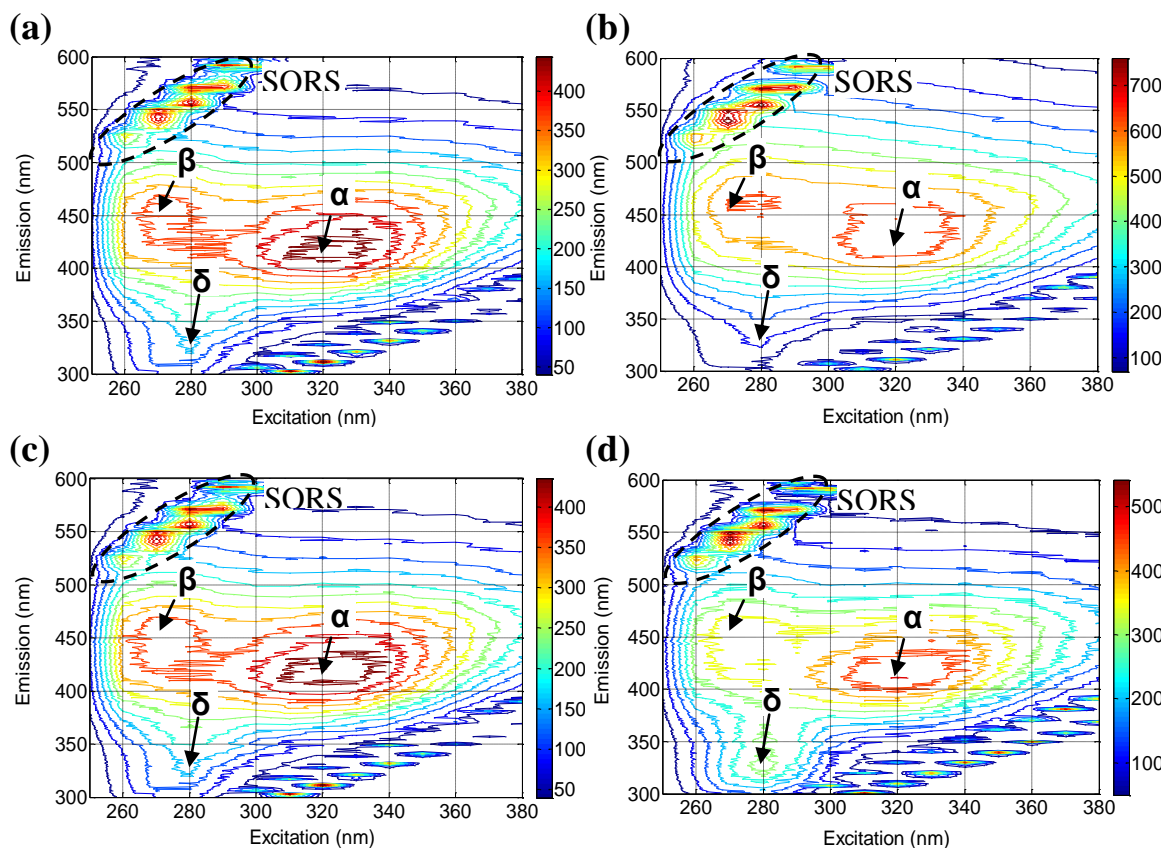


Figure 6-3 Fluorescence EEMs of (a) GRW and GRW spiked with (b) AHA, (c) AA and (d) AL. Second order Rayleigh scattering (SORS) regions in the fluorescence EEMs are indicated using dashed-circles.

In the fluorescence EEM of GRW, the fulvic acid-like NOM surrogate (Figure 6-3a) showed a peak (α) at Ex/Em \sim 320 nm/420 nm, which corresponds to the range reported for HS (Coble *et al.*, 1990) and specifically for fulvic acid-like NOM (Sierra *et al.*, 2005). Additional peaks were also observed including a secondary peak (β) that appears to be present in the form of a shoulder to peak (α) around Ex/Em = 270 nm/460 nm (Figure 6-3a) which is related to the presence of more humic-acid like NOM (Sierra *et al.*, 2005). The deviations of the fluorescence EEM contours seen in the region (Ex/Em: 280 nm/330 nm) indicated by δ is considered to be due to the presence of protein-like NOM in the water. The existence of a fluorescence EEM peak around the same region (δ) has been previously observed for protein-like substances (Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003). The protein-like NOM peak in the δ region is not clearly visible due to the very low concentration levels of protein-like substances present in GRW. Figure 6-3b provides the fluorescence EEM of a mixture in which 3 mg-DOC/L of AHA was spiked into 3.25 mg-DOC/L GRW. As a result of spiking of AHA the intensities at peaks α , β and δ increased. Most importantly, unlike in the fluorescence EEM of GRW (Figure 6-3a), peak β appeared as a complete EEM peak (i.e. not a shoulder to peak α) demonstrating that the addition of AHA was responsible for the appearance of a clear humic-acid like peak. This supports the suitability of AHA as a surrogate for humic acid-like NOM in natural waters. Addition of AA (1 mg-DOC/L) into GRW did not significantly change the fluorescence spectral properties of GRW as seen in Figure 6-3c. This is expected, since AA – a surrogate polysaccharide-like NOM, does not fluoresce (SI Figure S - 6-2). When AL (1 mg-DOC/L) was added to GRW, the peak (δ) appeared as a distinct peak (Figure 6-3d) supporting its use as a surrogate for protein-like NOM in water. In addition to these fluorescence regions, the second order Rayleigh

scattering (SORS) regions were also significantly affected due to the addition of AHA, AL and AA into GRW as discussed later.

Figure 6-3a–d illustrate that the changes in the fluorescence EEMs of GRW that occurred after the addition of AHA and AL are due to inherent fluorescent spectral properties of AHA and AL. These changes which have unique signatures served as the basis for identifying the changes that occur in the humic acid-, fulvic acid- and protein-like NOM in water. The addition of AA into GRW however did not show significantly visible spectral changes. However due to the expected potential interactions between polysaccharide-like matter (Imeson *et al.*, 1977; Zhao *et al.*, 2009), subtle changes in the spectra were expected to occur. Therefore, a systematic ANOVA statistical study was performed to understand how the fluorescence EEM intensities are affected by changes in the different NOM fraction surrogates at different concentrations and to determine whether there were significant interactions between the different NOM surrogates using a factorial experimental design (Table 6-1). The main findings of this analysis are summarized here whereas a more detailed version of the results can be found in SI Table S - 6-1.

1. Both peak (α) and peak (β) intensities are significantly affected by the presence of AHA and GRW as indicated by the very low P-value (< 0.01) (Table 6-1).
2. In addition to these main effects, the peak (β) intensity is also significantly affected by the presence of interactions between protein- and polysaccharide-like NOM surrogate as denoted by AL x AA in Table 6-1 (P-value < 0.05).
3. Peak (δ), is significantly affected by the presence of AL, GRW and the interactions between AL and AA.

4. Presence of AA alone does not affect fluorescence intensities of peaks α , β and δ .
5. SORS peaks on the other hand are affected by the presence of all four NOM surrogates. The interactions between AL and AA are also significant in the SORS region (Table 6-1).
6. Non-linear behaviour that is present in the fluorescence EEM is mainly due to the AL – AA interactions that were found to be significant (Table 6-1).

Table 6-1 Results of the ANOVA that illustrates the effect of NOM fraction surrogates on fluorescence EEM intensities.

EEM peak/region	A	β	δ	Second order Rayleigh scattering (SORS) peaks
Ex (nm)/Em (nm):	320/420	270/460	280/330	
Main effects				
Aldrich humic acid (AHA)	***	***		***
α -lactalbumin (AL)			***	***/*
alginic acid (AA)				***/*
Grand River water (GRW)	***	***	***	***
Interactions				
AL x AA		*	*	***

*** - highly significant: P-value < 0.01; * - significant: 0.01 < P-value < 0.05

6.3.3 Model Predictions

The spectral differences seen in the fluorescence EEMs of the mixtures at different concentrations of AHA, AL, AA and GRW were used to develop a calibration model to estimate the DOC content of samples containing different NOM fractions (Section 6.2.8).

PLS analysis resulted in a regression model (i.e calibration model) with seven statistically significant LVs. This calibration model with seven LVs demonstrated a goodness of prediction (Q^2) by cross validation of about 98%, capturing 98.3% of the variance in the X_c matrix and 98.8% of the variance in the Y_c data matrix. The rest of the variability is assumed to be due to experimental error and instrument noise. The model prediction results demonstrate (Figure 6-4a - d) how the calibration model performed in estimating the individual DOC values of AHA, AL, AA and GRW for an independent validation dataset (matrix X_v).

Figure 6-4a – d shows comparisons between the model predictions and the individual measured DOC values of all the samples, with different combinations of AHA, AL, AA and GRW concentrations, in the validation sample set. These individual DOC measurements were obtained based on the original DOC measurements of GRW, AHA, AL and AA solutions used for sample preparations (Section 6.2.5). All predictions show very good agreement with the individual measured DOC concentrations demonstrating that the fluorescence-based calibration model was suitable for identifying the individual changes in the concentration levels of AHA, AL, AA and GRW in the validation dataset. The maximum prediction error (MPE) for AHA was about 0.21 mg-DOC/L which corresponds to approximately 7% of the total variation of the AHA concentration. The MPE values recorded for AL, AA and GRW were approximately 13%, 15% and 4%, respectively of the total variation of their corresponding concentrations. These maximum errors only occurred for a small percentage of the cases tested. Prediction errors for the fraction concentrations in most

instances did not exceed 3%, 5%, 10% and 10% and of the total variation for GRW, AHA, AL and AA, respectively.

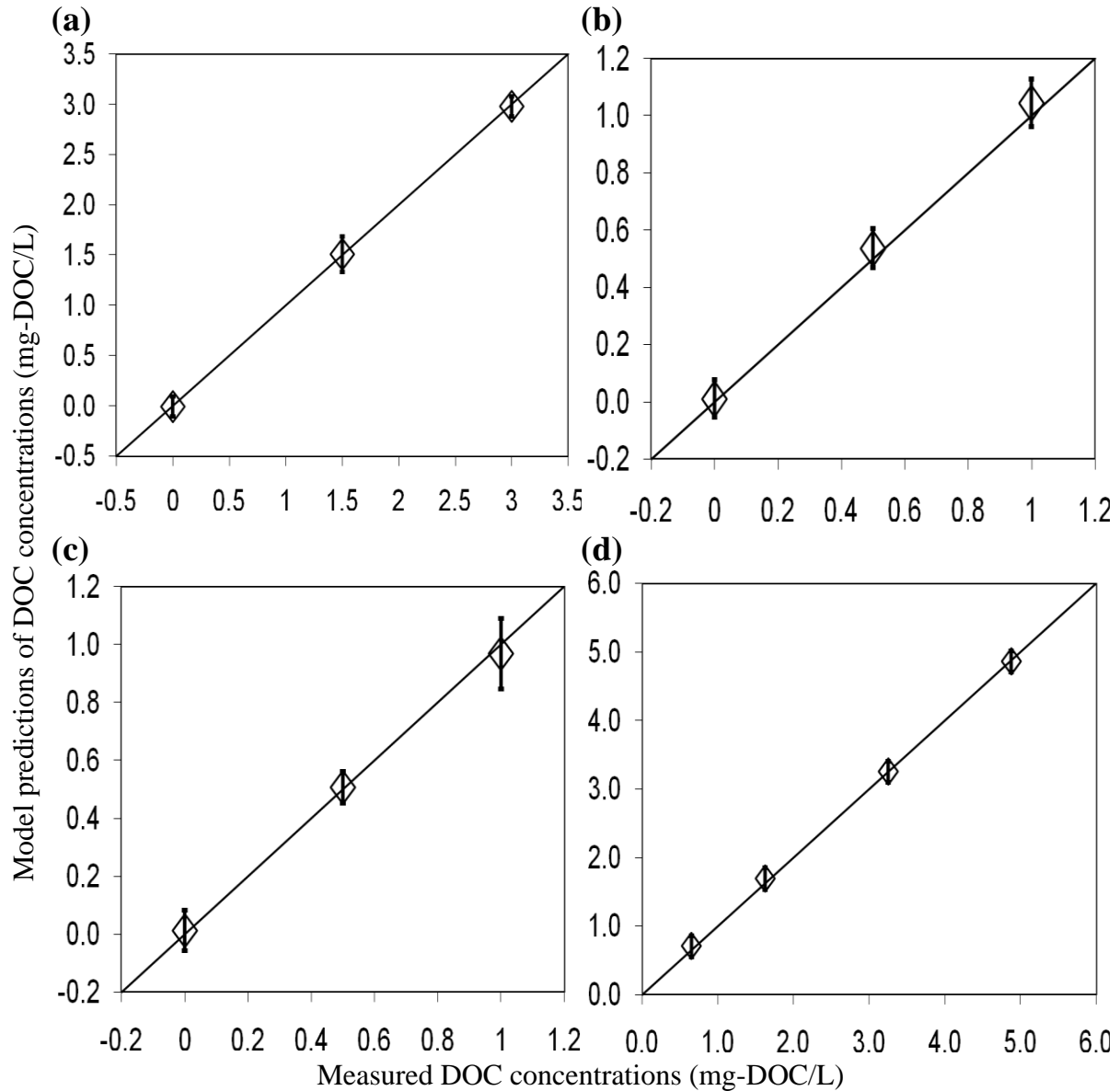


Figure 6-4 Comparisons between the model predictions and the individual measured DOC values of (a) AHA, (b) AL (c) AA and (d) GRW in water samples in the validation sample set. This validation sample set contains different concentration combinations of AHA, AL, AA and GRW. The error bars represent the maximum error at 95% confidence for the predictions at each predicted concentration level.

6.3.4 Significance of the Approach for Identification of Major NOM Fractions

Conventional fluorescence-based NOM characterization techniques such as fluorescence peak-picking or single-scan methods are not able to effectively capture the non-linear behaviour that was observed in the fluorescence EEM intensities (Table 6-1). Since the spectral variables in fluorescence EEMs are highly correlated, conventional regression methods such as multivariate least squares lack the ability to identify and distinguish the changes that occur for different NOM fractions in water. The PLS analysis of full fluorescence EEMs addressed these issues effectively and thus the major NOM fractions could be characterized as demonstrated by the predictions (Figure 6-4a – d). In addition to the NOM fractions that fluoresce, this approach was also suitable for characterizing non-fluorescent polysaccharide-like fractions by utilizing the changes in fluorescence EEM regions that resulted due to interactions between protein-like and polysaccharide-like NOM fractions.

The analysis (Section 6.3.2) demonstrated that the presence of AA in the water significantly changed the fluorescence EEM intensities in water samples (Table 6-1). The calibration model was able to capture these changes and relate them to the changes in AA concentrations. This could be confirmed by examining the sensitivity spectra of the calibration model. Sensitivity spectra (Boehl *et al.*, 2003) of the PLS-based calibration models are used to understand the areas in the fluorescence EEMs that are relevant for predicting the AHA, AL, AA and GRW concentrations. For each predictor variable (i.e. DOC content of AHA, AL, AA and GRW), a sensitivity spectrum, consisting of the scaled regression coefficients that correspond to each wavelength combination in the model, was

generated. When the sensitivity spectra of predictor variables AHA, AL, and GRW were examined, the peak of each sensitivity spectrum demonstrated a close resemblance to their corresponding fluorescence EEM peak (SI Figure S - 6-3a, b and d). The sensitivity spectrum of the predictor variable AA contained peaks in the regions corresponding to SORS region, peak (β) and peak (δ) of the fluorescence EEMs (SI Figure S - 6-3c). These EEM regions were found to be significantly affected by the presence of AA in the sample (Section 6.3.2 and Table 6-1) confirming that the model was able to capture changes that occurred in the fluorescence EEMs due to the presence of AA.

6.3.5 Characterizing the Interactions between Protein-like and Polysaccharide-like NOM Fractions

The PLS-based calibration model performed well in predicting all four surrogate NOM fractions even with the non-linear behaviour displayed for the fluorescence EEM intensities. It is reasonable to assume that this non-linear behaviour, which was manifested in terms of interactions between AL and AA, was captured in the calibration model. This means that if the interaction between AL and AA (i.e. AL x AA; x- indicates multiplication of AL and AA concentrations) is included as an additional output variable during the PLS model calibration, the resulting model should be able to predict changes of the interaction between AL and AA as it did for individual surrogate NOM fractions. When this was implemented by including “AL x AA” as a new predictor variable, in addition to the predictor variables discussed earlier in the PLS calibration, the new calibration model was able to predict the interaction between AL and AA with good accuracy (Figure 6-5). The MPE was approximately 12% of the total variation in the scale examined with MPE not exceeding 10% in most instances.

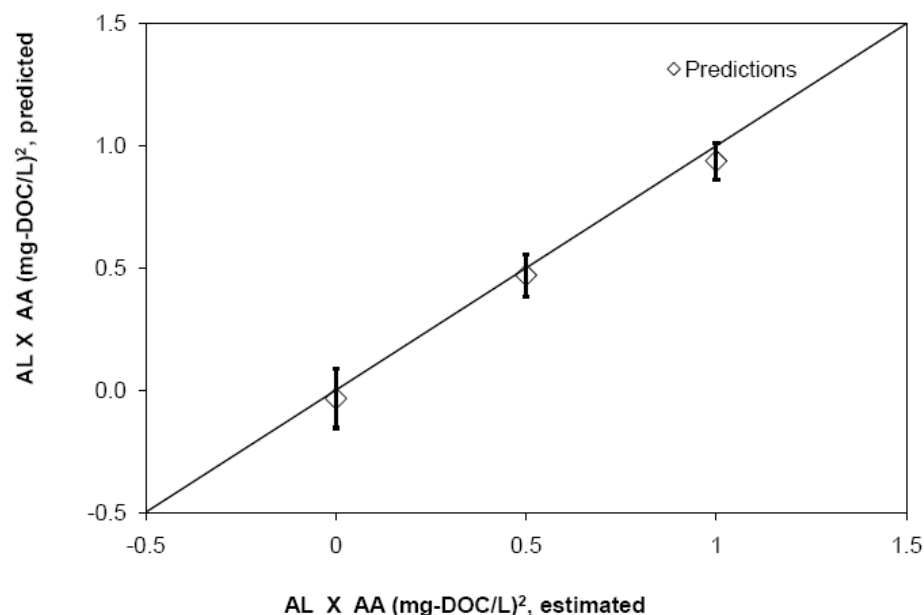


Figure 6-5 Comparisons between the model predictions and the estimated values for interaction between AL and AA based on measured DOC. The error bars represent the maximum error 95% confidence for the predictions at each predicted concentration level.

6.3.6 Potential for Process Monitoring and Other Applications

This study demonstrated that the proposed fluorescence EEM-based NOM characterization was able to identify changes in individual surrogate NOM fractions with good accuracy. Since this approach was developed based on surrogate NOM fractions that have similar properties, as confirmed by LC-OCD and fluorescence analyses, to the NOM fractions in natural water, it is proposed that this approach could serve as a rapid NOM characterization method for identifying the changes in different NOM fractions that may occur during different stages in a drinking water treatment process. This approach could be used to monitor water treatment processes where very low NOM concentration levels are observed such as membrane processes and disinfection stages or to identify the contaminations that may occur in drinking water distribution systems. In contrast, most other reported NOM

characterization techniques require pre-concentration steps prior to the analysis of water with low NOM concentrations, thereby increasing the chances for higher measurement noise.

The capability of this approach in characterization of the interaction between polysaccharide-like and protein-like NOM present in water, is very promising for the monitoring of membrane fouling in drinking water treatment applications. Interplay between polysaccharide-like and protein-like NOM could lead to different membrane fouling behaviour as indicated in many studies (Zularisam *et al.*, 2007; Zheng *et al.*, 2009) and therefore the characterization of such interactions would provide an opportunity to enhance our fundamental understanding on how these interactions relate to membrane fouling.

The fluorescence EEMs obtained during this study were obtained using off-line measurements, and the signal acquisition time for each EEM was about 5 min. Therefore, this approach could be readily used in rapid off-line analysis at low cost. Further studies directed at minimizing the number of intensity readings needed to be recorded without compromising the accuracy of the approach would be helpful in reducing signal acquisition time. An approach is proposed in Section 6.5.1 to this end.

6.4 Conclusions

This study focused on developing a fluorescence-based calibration model that could identify the changes in the four major NOM fractions: fulvic acid-, humic acid-, protein- and polysaccharide-like, in natural water. The following conclusions can be drawn:

1. The fluorescence-based calibration model predictions demonstrated a very good agreement with the measured DOC concentrations based on measured DOC for the surrogate NOM fractions used in this study.
2. The maximum prediction error (MPE) recorded for AHA and GRW was less than 7%, while the MPE of AL and AA were higher, reaching 13 % and 15%, respectively.
3. Factorial ANOVA performed on the fluorescence EEM intensities of surrogate water samples revealed that fluorescence intensities were affected not only by the presence of fluorescing NOM fractions but also by the presence of non-fluorescing polysaccharide-like NOM in water.
4. Significant non-linear interaction was seen to be present in the fluorescence EEMs of the surrogate water samples. This non-linearity appeared to be linked to interactions between polysaccharide-like and protein-like NOM present in water.
5. The ability to rapidly characterize polysaccharide-like NOM and interactions between polysaccharide-like and protein-like NOM is novel and possibly of considerable practical value.
6. The benefit of this approach is related to the capability of the calibration model to capture significant spectral changes that occurred in the fluorescence EEMs due to the presence of AA – a polysaccharide-like NOM.

6.5 Supplementary Information

Step-1

Rearrangement of fluorescence EEM intensity values to a row

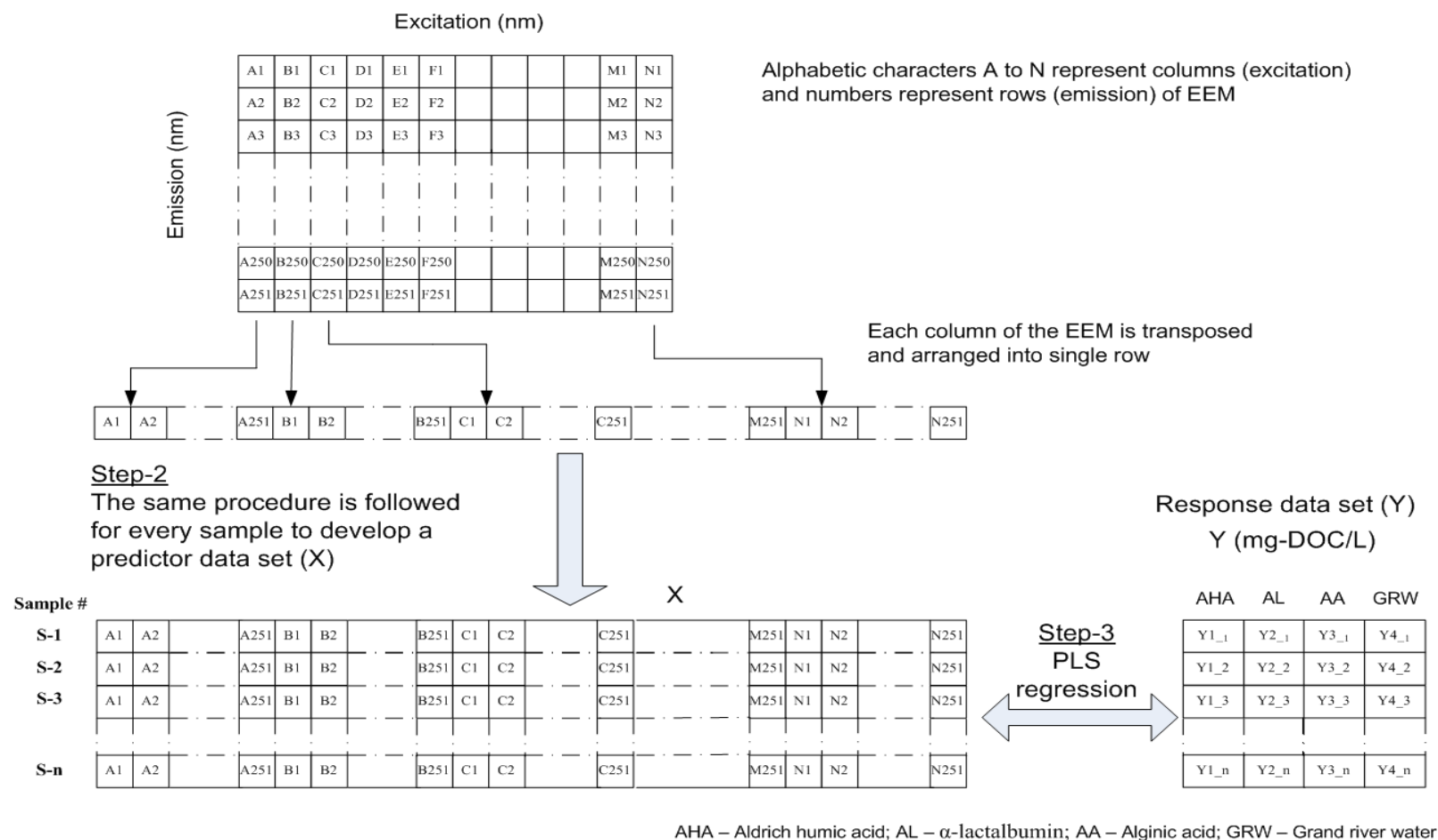


Figure S - 6-1 Steps involved in generating the fluorescence-based calibration model.

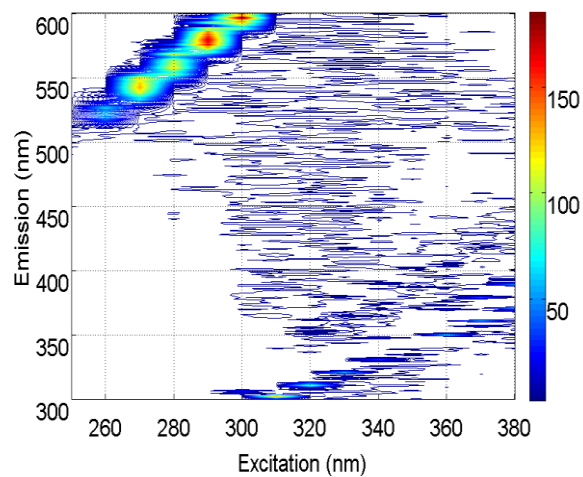


Figure S - 6-2 Fluorescence EEM of alginic acid (20 mg/L), which demonstrates that alginic acid does not fluoresce.

Table S - 6-1 Results of the ANOVA that illustrates the effect of surrogate NOM fractions on fluorescence EEM intensities.

EEM peak/region	α	β	δ	Second-order scattering					
Ex (nm)/Em (nm):	320/420	270/460	280/330	250/500	260/520	270/540	280/560	290/580	300/600
Main effects									
Aldrich humic acid (AHA)	***	***		***	***	***	***		
α -lactalbumin (AL)			***	*	***	***	***	*	*
alginic acid (AA)				*	*	*	*	***	***
Grand River water (GRW)	***	***	***	***	*	***	*	*	*
Interactions									
AL x AA		*	*	***	***	***	*	*	*
AHA x AL x AA				*					

*** - highly significant: P-value < 0.01; * - significant: 0.01 < P-value < 0.05

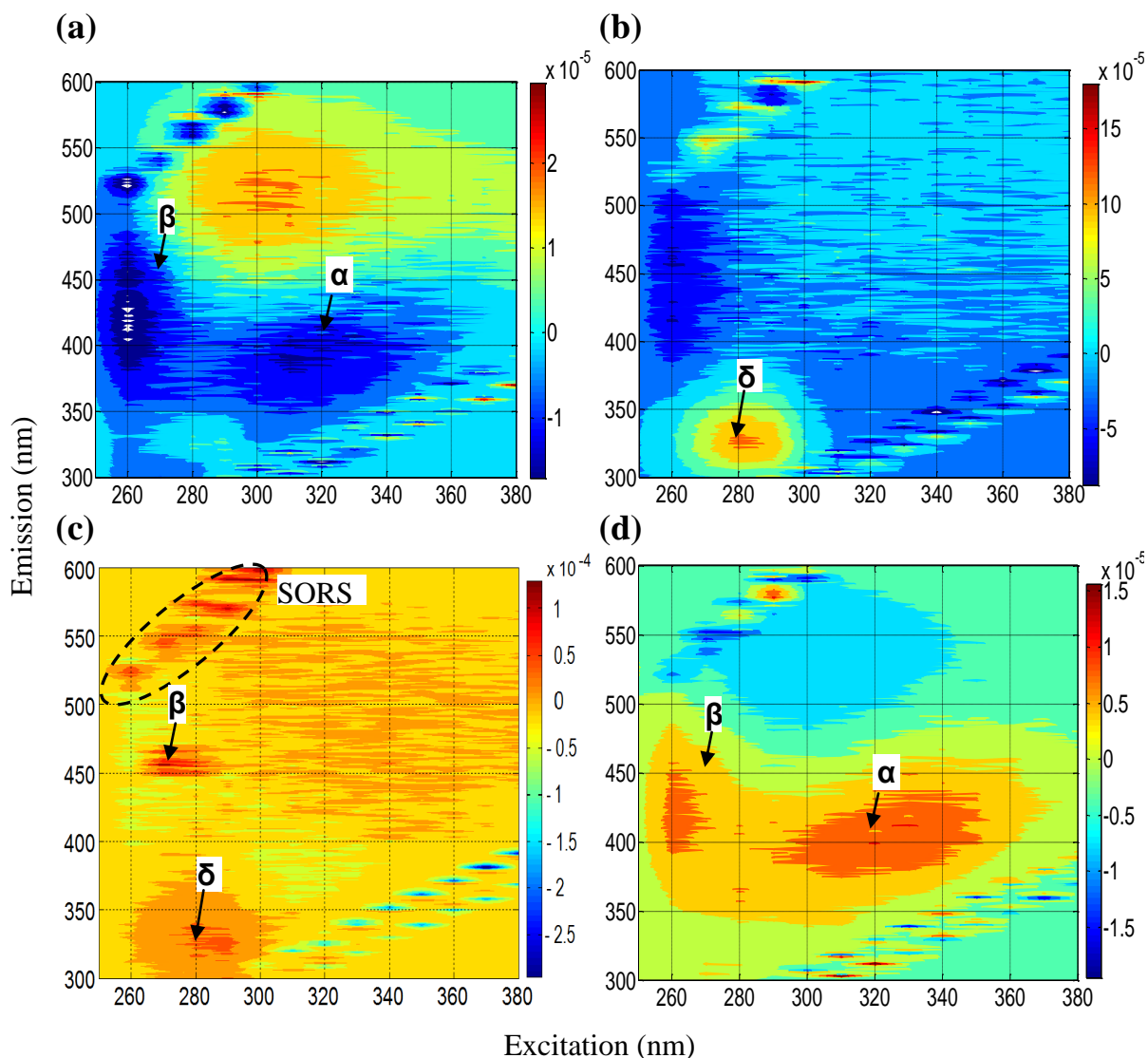


Figure S - 6-3 Sensitivity spectra of the PLS-based calibration model that illustrate areas in the fluorescence EEMs which are relevant for predicting the AHA, AL, AA and GRW concentrations. Sensitivity spectra of predictor variables (a) AHA, (b) AL, (c) AA and (d) GRW are closely linked with their corresponding fluorescence EEM peaks.

6.5.1 Optimum Scanning Path Approach for Fluorescence Signal Acquisition

This section proposes a fluorescence scanning approach to minimize the fluorescence signal acquisition time as discussed in Section 6.3.6. Figure S - 6-4 illustrates a fluorescence scanning path identified based on the sensitivity spectra of predictor variables AHA, AL, AA and GRW. The rationale for using this fluorescence scanning path was that it covered the

regions containing sensitivity spectra peaks (SI Figure S - 6-3a, b, c and d) of these predictor variables. The sensitivity peaks consist of scaled regression coefficients of higher magnitude compared to other wavelength combinations in sensitivity spectra. Therefore, it is reasonable to conclude that fluorescence excitation-emission wavelength combinations corresponding to these sensitivity peak regions have a larger contribution to the model predictions. The number of fluorescence excitation-emission wavelength combinations used in this proposed optimum scanning approach was derived by trial and error that involved considering wavelength combinations related to sensitivity peak regions. This resulted in 266 optimum excitation-emission wavelength combinations that are indicated by the solid lines in the optimum scanning path presented in Figure S - 6-4.

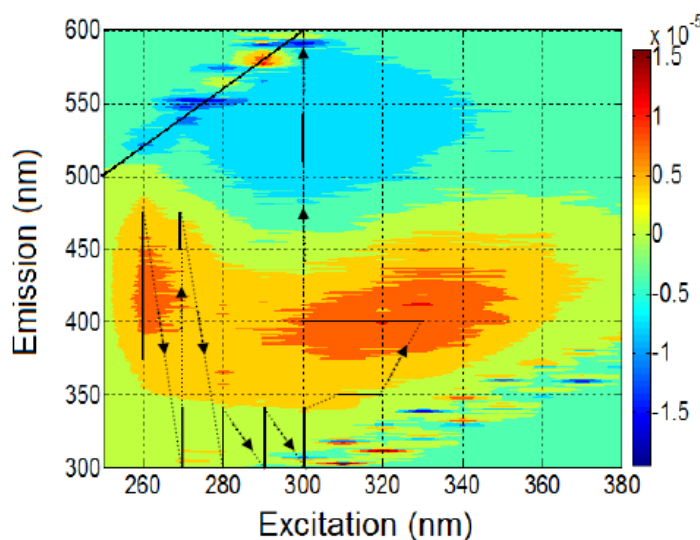


Figure S - 6-4 Optimum scanning path that was used to minimize the fluorescence acquisition time. Solid lines indicate the optimum excitation-emission wavelength combinations scanned.

The fluorescence intensity data obtained from these 266 excitation-emission wavelength combinations (i.e. optimum scanning path) for all the samples in the calibration set were used

to generate a PLS-based calibration model as explained earlier (Section 6.2.8). Fluorescence intensity data obtained from the same approach for the validation sample set were used to obtain model predictions. Compared to the fluorescence EEM approach used in this study, the model predictions obtained with the proposed optimum scanning approach were less accurate. The maximum prediction errors recorded for AHA, AL, AA and GRW were ~ 9 %, 17.2 %, 19.3 % and 7.1 %, respectively, of the total variation in the scale examined. Although this indicates that it is possible to minimize the fluorescence data acquisition time by using a smaller number of excitation-emission wavelength combinations, this will however diminish the model prediction accuracy.

CHAPTER 7

Assessing Nanofiltration Fouling in Drinking Water Treatment Using Fluorescence Fingerprinting and LC-OCD Analyses^{*}

Overview

The natural organic matter (NOM) components causing fouling of nanofiltration membranes used in drinking water applications consists in a complex mixture of humic and fulvic acids, proteins, and carbohydrates of various molecular size and functional groups. Understanding the characteristics of NOM fractions such as humic substances (HS) and biopolymers (proteins and polysaccharides) as foulants is of paramount importance to develop fouling control strategies. Fluorescence spectroscopy is becoming an increasingly popular method for characterizing NOM and shows good potential for on-line monitoring, as minimal sample pre-treatment and preparation is required, high instrumental sensitivity is available and the technique is non-destructive in nature. In this research an innovative approach involving both fluorescence and LC-OCD analyses is used to identify and to characterize organic membrane foulant.

^{*} Peiris, B. R. H., Hallé, C., Haberkamp, J., Legge, R. L., Peldszus, S., Moresoli, C., Budman, H., Amy, G., Jekel, M., Huck, P. M. 2008. Assessing nanofiltration fouling in drinking water treatment using fluorescence fingerprinting and LC-OCD analyses. *Water Sci. and Technol.: Water Supply* **8** (4), 459-465

Keywords: drinking water treatment; fluorescence spectroscopy; LC-OCD; membrane foulants; nanofiltration

7.1 Introduction

The use of nanofiltration (NF) membranes to produce drinking water from surface water sources is increasing; however, membrane fouling is the main limitation to this technology. Fouling is caused by the deposition and accumulation of material on the surface of the membrane or within the internal structure of the membrane. Fouling increases the hydraulic resistance and operational costs.

Different types of fouling may occur during membrane processes; the main types are organic, inorganic, colloidal and particulate, and biofouling. In drinking water NF applications, natural organic matter (NOM) is a major membrane foulant. In general, NOM consists of a complex mixture of humic and fulvic acids, proteins, and carbohydrates of various molecular size and functional groups. The nature of the NOM in water can vary significantly depending on the location, climate and hydrological conditions. Effluent organic matter from upstream municipal wastewater treatment plants can also be an important source of organic matter.

Understanding the characteristics of NOM fractions such as humic substances (HS) and biopolymers (proteins and polysaccharides) as foulants is of paramount importance to develop fouling control strategies. Reported methods for the characterization of NOM have mostly focused on traditional analyses such as light absorptivity, dissolved organic carbon

(DOC) concentration and aromaticity due to the difficulty of detailed structural analysis (Her *et al.*, 2003). More advanced methods for analysing the structural characteristics of NOM such as nuclear magnetic resonance (^{13}C NMR) spectroscopy, infrared (IR) spectroscopy, differential thermal analysis, modulated differential scanning calorimetry, pyrolysis-gas chromatography/mass spectrometry and Fourier transform infrared spectroscopy (FTIR) have also been reported (Wilson *et al.*, 1999; Smeulders *et al.*, 2000; Lee *et al.*, 2006; Gray *et al.*, 2007). Most of these characterization methods, however, require concentration and/or fractionation of NOM to accommodate challenges associated with measuring the low NOM concentrations found in raw water (Croué, 2004).

High performance size exclusion chromatography (HPSEC) can be used to determine the molecular size distribution of the dissolved NOM. Moreover, HPSEC coupled with UV and DOC detectors can be used to determine the specific nature of the NOM (Her *et al.*, 2008). HPSEC has various advantages including small injection volumes and minimal sample pre-treatment but the analysis is expensive, requires sophisticated equipment and is not suitable for on-line application. In this context, fluorescence spectroscopy is becoming an increasingly popular method for characterizing NOM (Her *et al.*, 2003) and shows good potential for on-line monitoring, as minimal sample pre-treatment and preparation is required, high instrumental sensitivity is available and the technique is non-destructive in nature.

This paper demonstrates the feasibility of using fluorescence spectroscopy to detect the differences in permeates from two NF membranes that exhibited different fouling rates and

how the differences in fluorescence spectra can be helpful in interpreting different rejection characteristics of NF membranes. LC-OCD analyses were used as an independent method to complement and validate the findings.

7.2 Methods

7.2.1 Water Samples and NF Set-up

Grand River water (Southwestern Ontario, Canada) was used as feed water during this study. The river is impacted by urban runoff, agricultural activity and wastewater effluent. Characteristics of Grand River water (GRW) during the experimental period (August to September 2007) are presented in Table 7-1 (column 2).

Table 7-1 Characteristics of Grand River water for the period August to September 2007.

Parameters	Grand River water	Biofilter Effluent	XN45 permeate	TS80 permeate
Temperature (°C)	17 - 22	20	26	26
pH	7.50 - 8.00	8.48	7.72	8.5
Turbidity (NTU)	2.78 - 27.2	0.53	0.45	0.11
TOC (mg/L)	6.8 - 8.1	7.3	1.3	0.3
DOC (mg/L)	6.3 - 8.0	6.7	0.8	0.4
SUVA (L/mg-M)	2.4 - 4.0	3.0	0.7	1.2
Conductivity (mS/cm)	540 - 588	607	491	94
Ca ²⁺ (mg/L)	54 - 62	60	46	12
Mg ²⁺ (mg/L)	19 - 23	18	16	0

GRW was first filtered through a roughing filter to lower the turbidity level of raw water prior to biofiltration. The roughing filter was constructed of PVC piping and the diameter of the column was designed to provide a $d_{\text{column}}/d_{\text{media}}$ of 10. The column was 1.5 m in height

with a diameter of 0.2 m. The column was filled with three layers of gravel with decreasing media sizes from bottom to top. The roughing filter was operated in an upflow mode at a rate of 1.1 m/h using a constant head tank. The biofilter was operated in a downflow mode fed by the roughing filter effluent at a flow rate of 5 m/h. It consisted of a dual media filter (anthracite and sand) over a gravel support layer. The empty bed contact time of the filter was 14 minutes. The biofilter underwent an acclimation period of two months prior to the experiments described here.

Nanofiltration experiments were performed with a bench scale module (GE SEPA™ CFII) using flat sheet membranes as illustrated in Figure 7-1. Two different flat sheet NF membranes (XN45 and TS80) from TriSep Corporation (California, USA) were used for this study. The active layer of both NF membranes is made of polyamide and the molecular weight cut off (MWCO) provided by the manufacturer was 200 Da for both membranes. The hydrophobicity is determined by contact angle measurement and a large contact angle is representative of a hydrophobic surface. The contact angle of both TS80 and XN45 is $57^{\circ} \pm 1^{\circ}$. The average roughness of the active layer was 8.8 nm and 21 nm for TS80 and XN45, respectively as measured by atomic force microscopy (AFM, diMultiMode, Veeco, US). The nominal surface area of both membranes was 0.0140 m^2 . The NF experiments with the XN45 and TS80 membranes were operated at a constant pressure of 8.2 bar and 12.4 bar, respectively. The pure water permeability of XN45 and TS80 were 10.4 LMH/bar and 10.0 LMH/bar, respectively. The initial recovery was 2% for both membranes. The initial permeate flux of XN45 and TS80 membranes was 85.7 LMH and 124.2 LMH. Prior to the experiments, the virgin membranes were compacted using de-ionized (DI) water until stable

permeate flow was achieved. Following the compaction period the DI water was drained from the feed tank and 20 L of feed water (biofilter effluent) was introduced into the tank. Throughout the experiments the temperature was kept constant at $25^{\circ}\text{C} \pm 2^{\circ}\text{C}$ through the use of a chiller. The experiments were performed in a recycle mode; both concentrate and permeate were returned to the feed tank. The duration of the experiment varied between 72h and 144h.

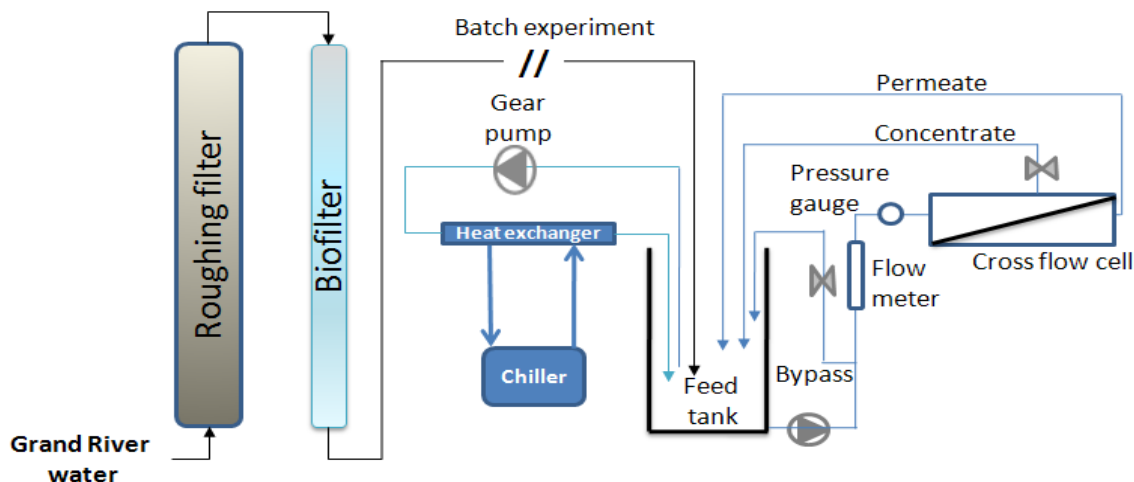


Figure 7-1 Schematic of the nanofiltration experimental set-up.

7.2.2 Analytical Methods

The concentration of total organic carbon (TOC) and dissolved organic carbon (DOC) were measured using an OI-Analytical TOC analyser (Model 1010, College Station, TX) with a wet-oxidation method as described in Standard Methods (2005) 5310D. Turbidity of the water samples was measured using a turbidity meter (Hach 2100P) following the Standard Method (2005) 2130. Conductivity was measured using a conductivity meter (Hach 44600) following the Standard Method (2005) 2510. Size exclusion chromatography with continuous organic carbon detection (LC-OCD) was performed with a DOC-Labor Dr. Huber

(Karlsruhe/Germany) system (Huber and Frimmel, 1992). The system uses a Toyopearl HW-50S SEC column (Tosoh Bioscience, Tokyo/Japan).

7.2.3 Fluorescence Analysis

The fluorescence excitation/emission matrix (EEM) of each sample was collected using a Varian Cary Eclipse Fluorescence Spectrofluorometer (Palo Alto, CA) by scanning 301 individual emission spectra (300 - 600 nm) at sequential 10 nm increments of excitation wavelength between 250 nm and 380 nm. UV-grade polymethylmethacrylate cuvettes with four optical windows were used in the analyses. The instrument parameters (photomultiplier tube (PMT) voltage = 800 V, scan rate = 600 nm/min and excitation/emission slit width = 10 nm each) were maintained during the fluorescence signal acquisition. These parameter settings were identified as optimum instrument settings for obtaining reproducible fluorescence signals, especially for the low concentrations seen in NF permeates, in a separate study. To eliminate water Raman scattering and to reduce other background noise, fluorescence spectra for Milli-Q (Millipore) water, obtained under the same conditions, were subtracted from all fluorescence spectra. The temperature of all water samples was maintained at room temperature (25°C) during the analyses. Data processing was performed using MATLAB 7.3.0 software (The Mathworks Inc., Natick, MA).

7.3 Results and Discussion

The fluorescence EEM of untreated GRW (RW) shows a peak (α) at Ex/Em = 320nm/415nm (Figure 7-2a), which corresponds to the range reported for HS (Sierra *et al.*, 2005). The

presence of HS in GRW can also be independently confirmed by examining the LC-OCD spectra of the same water sample (Figure 7-2b). LC-OCD spectra of RW also reveal that HS is the most plentiful fraction of NOM in GRW. In addition to the primary peak (α), another secondary peak (β) which also corresponds to HS (Sierra *et al.*, 2005) appears to be present in the form of a shoulder at Ex/Em = 270nm/450nm (Figure 7-2a).

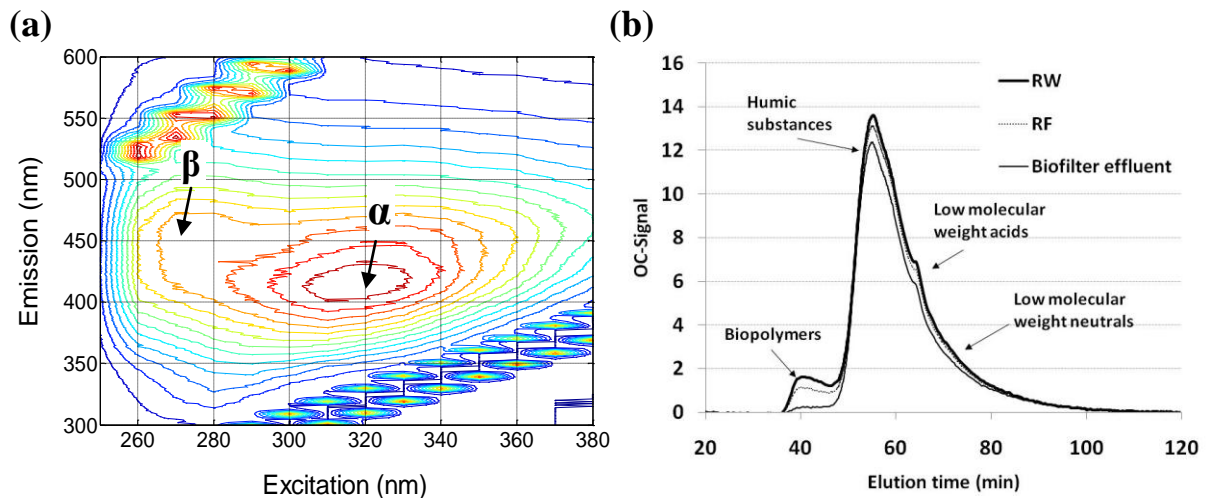


Figure 7-2 Fluorescence EEMs of (a) GRW (RW) compared with (b) corresponding LC-OCD spectra.

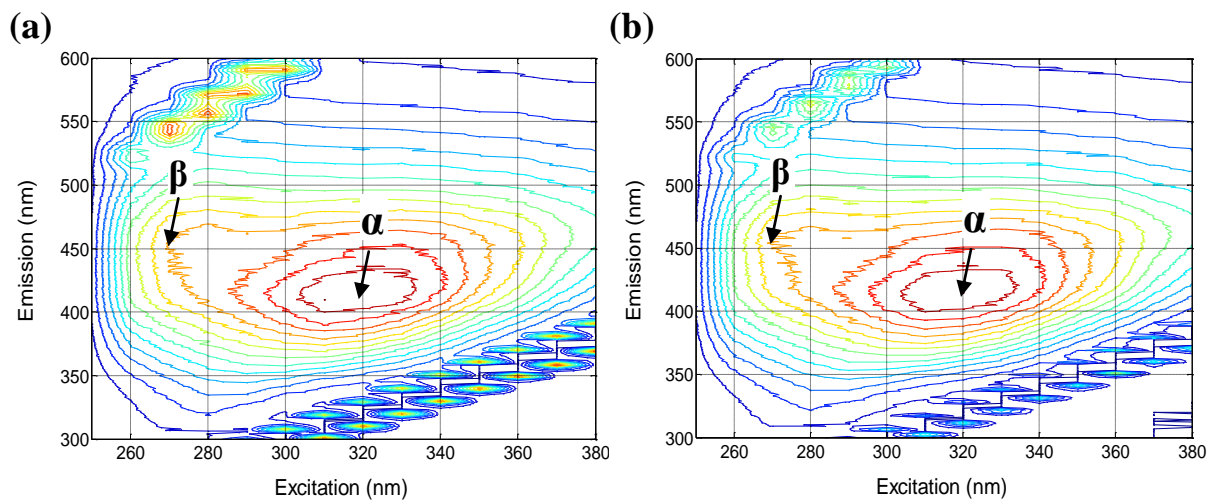


Figure 7-3 Fluorescence EEMs for (a) GRW filtered through the roughing filter (RF) and (b) biofilter effluent.

The fluorescence EEMs of GRW (RW), GRW filtered through the roughing filter (RF), and biofilter effluent (BF) appear to have similar spectral characteristics (Figure 7-3). Nevertheless, differences in the fluorescence intensities at peak (α), peak (β) and the peak position (Ex/Em: 280nm/330nm) corresponding to protein-like substances (Her *et al.*, 2003) of RW, RF, and BF suggest some removal of NOM by the roughing filter and biofilter (Table 7-2).

The removal of HS, in these stages, is substantially less on a percentage basis than the removal of biopolymers. The above observations can be confirmed through the comparison of both LC-OCD spectra and fluorescence peak intensities of corresponding NOM components in RW, RF and BF (Figure 7-2b and Table 7-2).

Table 7-2 Fluorescence intensities at reported peak positions for HS and protein-like substances. The values given within brackets (bold letters) are the intensities of peak (α').

	Intensity (a.u.)		
	Peak (α): Ex/Em: 320/415 (HS)	Peak (β): Ex/Em: 270/450 (HS)	Ex/Em: 280/330 (protein-like)
RW	904.0	694.6	201.9
RF-effluent	858.5	617.5	161.2
BF-effluent	849.4	606.3	152.6
NF-permeates			
XN45	88.5 (103.00 @ Ex/Em: 320/390)	54.7	44.2
TS80	11.6 (22.17 @ Ex/Em: 320/380)	20.9	45.6

The NF membranes XN45 and TS80 used in this study achieved significant levels of NOM removal, as would be expected. The fluorescence peak intensities of HS are reduced about tenfold or more, compared to the peak intensities of RW, depending on the type of NF membrane used. Fluorescence peak intensities that correspond to protein-like substances were, however, seen to be less affected on a percentage basis (Table 7-2). The peak (β) for HS, observed in the form of a shoulder in the EEMs of RW, RF, and BF, appears as a complete peak in the permeate streams (Figure 7-4c and d). The very weak, virtually nonexistent LC-OCD signals for the permeates of XN45 and TS80 NF membranes confirm the extent of NOM rejection (Figure 7-4a).

Even though both XN45 and TS80 membranes had similar MWCO and contact angle characteristics, the TS80 membrane demonstrated a substantially faster fouling rate with the same feed water. After 72h of operation the XN45 and TS80 membranes had a flux equivalent to 72% and 5% of the initial flux, respectively. Higher levels of TOC and DOC, conductivity, calcium and magnesium concentrations observed for the XN45 permeate indicate different retention properties (Table 7-1). Different fouling rates therefore suggest different fouling mechanisms and/or different foulant material as well as potential differences in membrane properties.

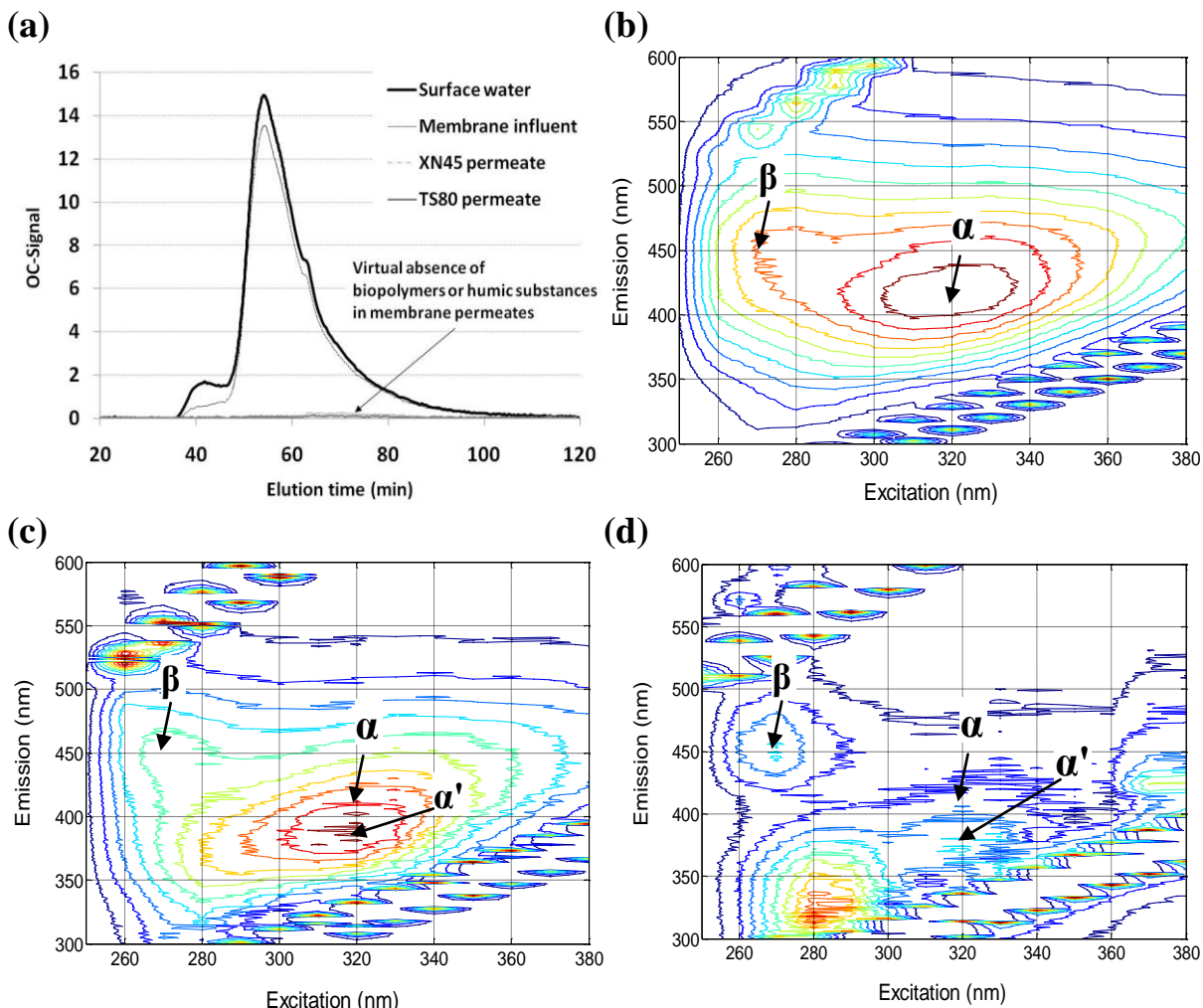


Figure 7-4 Fluorescence EEMs of the membrane feed (b) and the permeates of the NF membranes; XN45 (c) and TS80 (d) and the corresponding LC-OCD spectra (a).

Unlike the weak LC-OCD signals, fluorescence EEMs for the permeates of NF membranes, with very low levels of NOM (between 0.30 mg/L and 1.30 mg/L), contained many spectral details demonstrating the sensitivity of the technique at low concentrations. The LC-OCD signals of both XN45 and TS80 permeates do not show any significant sign of the presence of HS in the permeate (Figure 7-4a). The fluorescence EEMs of XN45 and TS80 permeates, on the other hand, display very different spectral characteristics. For example, the intensities of peak (α) and peak (β), which describe the extent of the presence of HS, are weaker in

TS80 permeates compared to the XN45 permeates (Table 7-2). Especially, the fluorescence EEM peak (α) appears to be the major peak in XN45 permeate, whereas the same peak is significantly weaker in TS80 permeate (Figure 7-4c and d, and Table 7-2). In addition, fluorescence peak (β) and the peak that corresponds to protein-like substances are more clearly noticeable in TS80 permeate than in XN45 permeate. These differences indicate that the composition of NOM in the permeate of the TS80 membrane is different than in the permeate of XN45. Therefore it is reasonable to assume that the different rejection properties of the two NF membranes, as indicated in Table 7-1, have likely contributed to the above differences in the permeates.

Moreover, EEMs of XN45 and TS80 permeates indicate a red shift (i.e., a peak shift towards a lower wavelength) in the position of peak (α) suggesting the absence of a certain fraction of HS. The new positions of peak (α) (peak- α') are found at Ex/Em: 320/390 and Ex/Em: 320/380 for XN45 and TS80 permeates, respectively (Figure 7-4c and d). This shift may have resulted from preferential rejection of a certain fraction of HS by both NF membranes. The difference in the positions of peak (α') between the XN45 and TS80 permeates also correlates with the different rejection (or retention) properties of the two NF membranes. These different rejection properties of the membranes appear to have played significantly different roles in the fouling of XN45 and TS80 membranes as discussed above.

Subsequently, in five more NF experiments which were limited to XN45 membranes due to the rapid fouling rates of the TS80 membranes, permeates with similar characteristics were produced even with measurable variation in raw water quality over a 3 month period. The

similarities in XN45 permeate quality were confirmed by three approaches (results not presented for reasons of space): (i) comparison of parameters such as TOC, conductivity, calcium and magnesium; (ii) absence of any significant LC-OCD spectra and (iii) almost identical fluorescence EEMs. These results suggest that the characterization of the permeates from the NF experiments as carried out in this study is reproducible and that fluorescence spectroscopy can be used to obtain reproducible information that may be useful for assessment of NF fouling. Note though that fluorescence EEM is limited to the detection of macromolecules containing a fluorophore (e.g., proteins and humic substances); whereas LC-OCD is capable of measuring organic based macromolecules such as polysaccharides in addition to proteins and humic substances. However, fluorescence EEM displayed a higher sensitivity than LC-OCD for the molecules it detected in this study.

7.4 Conclusions

This study demonstrated how fluorescence spectroscopy can be used to assess the NF fouling in drinking water treatment. The two NF membranes used in this study (XN45 and TS80) provided different levels of NOM removal. The LC-OCD spectra that were collected in parallel at different stages of the process also confirmed these findings. The fluorescence EEMs of XN45 and TS80 permeates indicated that the composition of NOM for the two NF membrane permeates is different. These differences, captured by fluorescence EEMs, appeared to have a correlation with the fouling behaviour observed. LC-OCD was however, comparatively less sensitive in capturing these differences. The LC-OCD chromatograms of both NF membranes showed essentially complete rejection of biopolymers and humic

substances, which offered little help in understanding the reasons for the different fouling rates observed. Further work should be directed at relating the fluorescence EEMs of NOM to the rejection properties of the membranes for predicting and mitigating membrane fouling.

CHAPTER 8

Identifying Fouling Events in a Membrane-based Drinking Water Treatment Process using Principal Component Analysis of Fluorescence Excitation-Emission Matrices^{*}

Overview

The identification of key foulants and the provision of early warning of high fouling events for drinking water treatment membrane processes is crucial for the development of effective countermeasures to membrane fouling, such as pretreatment. Principal foulants include organic, colloidal and particulate matter present in the membrane feed water. In this research, principal component analysis (PCA) of fluorescence excitation-emission matrices (EEMs) was identified as a viable tool for monitoring the performance of pre-treatment stages (in this case biological filtration), as well as ultrafiltration (UF) and nanofiltration (NF) membrane systems. In addition, fluorescence EEM-based principal component (PC) score plots, generated using the fluorescence EEMs obtained after just 1 hour of UF or NF operation, could be related to high fouling events likely caused by elevated levels of particulate/colloid-like material in the biofilter effluents. The fluorescence EEM-based PCA

^{*} Peiris, R. H., Hallé, C., Budman, H., Moresoli, C., Peldszus, S., Huck, P. M., Legge, R. L. 2010. Identifying fouling events in a membrane-based drinking water treatment process using principal component analysis of fluorescence excitation-emission matrices. *Water Res.* **44** (1), 185-194.

approach presented here is sensitive enough to be used at low organic carbon levels and has potential as an early detection method to identify high fouling events, allowing appropriate operational countermeasures to be taken.

Keywords: principal component analysis, fluorescence spectroscopy, membrane fouling, drinking water treatment; nanofiltration; ultrafiltration

8.1 Introduction

Membrane treatment of surface and ground water by means of ultrafiltration (UF) and nanofiltration (NF) is increasingly being used as an option for the production of drinking water. However, implementation of these membrane-based processes for drinking water treatment is often constrained due to fouling, which may be caused by organic, inorganic, colloidal and particulate matter. In drinking water UF and NF applications, natural organic matter (NOM) is considered to be the major membrane foulant (Saravia *et al.*, 2006; Jermann *et al.*, 2007). NOM consists of a complex mixture of humic and fulvic acids, proteins, and carbohydrates of various molecular size and functional groups (Her *et al.*, 2003).

Characterization of membrane foulant fractions in NOM such as humic substances (HS) and biopolymers (protein and polysaccharides) is indispensable for understanding membrane fouling and for the development of fouling control strategies (Amy, 2008). Application of fluorescence spectroscopy as a tool for characterizing NOM is well documented (Coble *et al.*, 1990; Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003; Sierra *et al.*, 2005; Hudson *et al.*, 2007; Henderson *et al.*, 2009). Compared to other available NOM characterization

techniques, this technique offers rapid and consistent analyses with high instrumental sensitivity (Peiris *et al.*, 2008).

In this study, the fluorescence excitation-emission matrix (EEM) analysis method was used for characterization of NOM and the associated fouling events in UF and NF-based drinking water treatment processes, as this method is able to capture specific fluorescence features that correspond to humic and protein-like materials in a single matrix in terms of fluorescence intensities. The light scattering regions captured in the fluorescence EEMs can also be used to provide information related to the particulate/colloidal matter present in water (Wyatt, 1993; Stramski *et al.*, 2005). In addition, unlike the fluorescence single scan approach (i.e. scanning only at the fluorescence peak location), the fluorescence EEM method provides a basis for capturing subtle changes in the fluorescence spectra of the water that may occur due to seasonal effects or other changes. The ability of this approach to characterize natural water NOM with a wide range of dissolved organic carbon (DOC) concentrations (i.e. 8.0 DOC-mg/L for raw water to 0.4 DOC-mg/L for NF permeate) without pre-dilution or pre-concentrations steps has also been demonstrated (Peiris *et al.*, 2008).

Most reported techniques examine fluorescence EEMs intensity data points at a few excitation-emission coordinate pairs (i.e. main peaks) from fluorescence spectra that may contain thousands of wavelength-dependent fluorescence intensity data points. These techniques lack the ability to capture the heterogeneity of the different NOM fractions in water. The importance of analyzing the full fluorescence EEMs as opposed to individual main peak positions has therefore been highlighted in several studies (Persson and Wedborg,

2001; Chen *et al.*, 2003; Stedmon *et al.*, 2003; Boehme *et al.*, 2004). Due to these reasons, full fluorescence EEMs of the water samples were analyzed in this study. Multivariate data analysis methods such as principal component analysis (PCA) (Persson and Wedborg, 2001; Boehme *et al.*, 2004) and parallel factor analysis (Stedmon *et al.*, 2003) have been used to analyze the full fluorescence EEMs to characterize water samples obtained from different sources/sampling locations. In contrast to these studies, the objective of the present study was to de-convolute the spectral information to identify major foulants present, and thereby to assess the performance of different feed water pre-treatment stages and the subsequent UF/NF stages. Since this objective could be satisfactorily met with PCA, this was the only data-mining technique used in this study. The application of this approach as a potential tool for early detection of high membrane fouling events is also described.

8.2 Materials and Methods

8.2.1 Feed Water and Pre-treatment

Water from the Grand River (Southwestern Ontario, Canada) was used as the feed water for UF and NF experiments conducted between August 2007 and August 2008. Typical Grand River water (GRW) quality parameters recorded during the experimental period are presented in Table 8-1.

Table 8-1 Grand River water quality parameters from August 2007 – August 2008.

Parameters	Grand River raw water
Temperature (°C)	1 - 23
pH	7.30 - 8.40
Turbidity (NTU)	1.45 - 67
DOC (mg/L)	5 - 9
Conductivity (mS/cm)	500 - 1200

Figure 8-1 demonstrates a process flow chart of the experimental set-up used in this study. GRW was first filtered through a roughing filter to lower the turbidity level of raw water (RW) prior to biofiltration. The roughing filter was operated in an up-flow mode at 1.1 m h^{-1} . More details about the roughing filter can be found in Peiris *et al.* (2008). The roughing filter effluent (RF) was then processed through one of the two parallel biofilters which consisted of dual media filters (i.e. anthracite and sand) over a support layer of gravel. The biofilters were operated in a down-flow mode at 5 m h^{-1} . The empty bed contact times (EBCT) of the two biofilters were 5 min (BF1) and 14 min (BF2), respectively. Further details on biofilter design are available (Hallé *et al.*, 2009). The effluents of BF1 (B1) and BF2 (B2) were then used as the feed for both UF and NF experiments. Table S - 8-1 and Table S - 8-2 under Supplementary Information (SI) summarizes when B1 and B2 were used as the membrane feed for different UF and NF experiments. These tables also identify which UF and NF experiments experienced high fouling events. The biofilters operated continuously, independently of whether the membranes units were in operation.

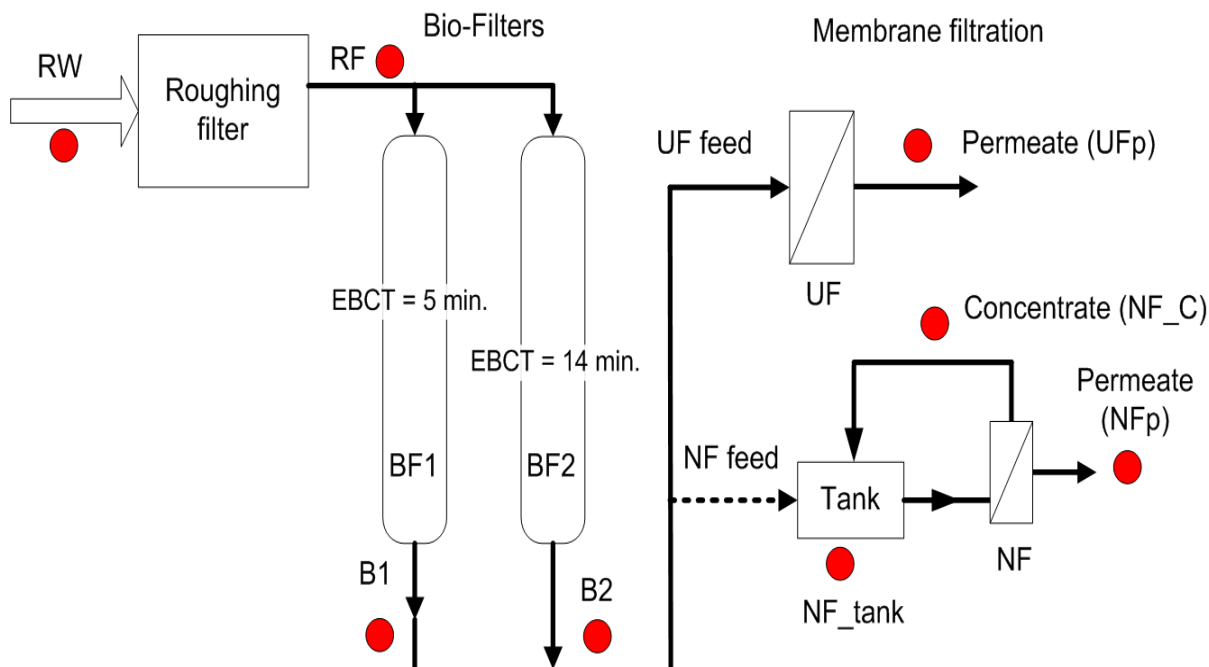


Figure 8-1 Pilot scale membrane filtration experimental set-up. Circles indicate the sampling points for water samples. The acronyms represent the following: BF1 - biofilter with empty bed contact time (EBCT) of 5 min; BF2 - biofilter with EBCT = 14 min; RW - raw GRW; RF - roughing filter effluent; B1- effluent of BF1; B2 – effluent of BF2; UFp - UF permeate; NF_C - concentrate of NF; NF_tank - water in the NF feed tank; NFp - NF permeate.

8.2.2 Pilot-scale Membrane Filtration Set-up

8.2.2.1 Ultrafiltration Membrane

A bench scale UF membrane module made of commercial hollow fibre membranes was used for this study (ZeeWeed – 1 by GE-Zenon, Oakville, Canada). The membrane consisted of PVDF and had a MWCO of 200 kDa. The membrane module had a surface area of 0.047 m², operated in outside-in mode and was mounted in a cylindrical holder of 1.6 L. The membrane was operated in a dead-end filtration mode at constant flux. The permeate flux was temperature adjusted to correspond to 57.5 LMH at 20°C, and the membrane was operated at a recovery of 94%. The four-step operation cycle was automated and consisted of: (1) permeation for 1 h, (2) back pulsing with air sparging for 20 sec, (3) draining 0.4 L from

the tank, and (4) filling the tank for 9 min. Each experiment was conducted for a 5 d period during which the UF unit was continuously fed by one of the biofilters. The trans-membrane pressure (TMP) was measured using pressure transducers. Under high fouling conditions (identified later), the TMP required to maintain the preset flux, exceeded the recommended operating range for the UF module and as a consequence, the permeate flux declined. A schematic with a more detailed description of the UF membrane set-up is available elsewhere (Hallé *et al.*, 2009).

8.2.2.2 Nanofiltration Membrane

NF experiments were performed using a bench scale module (GE SEPATM CFII). The system setup and operational conditions are described elsewhere (Peiris *et al.*, 2008). XN45 and TS80 flat sheet membranes from TriSep Corporation (California, USA) were used. The active layer of the membranes was polyamide and the MWCO provided by the manufacturer was 200 Da for both membranes. XN45 and TS80 are hydrophobic membranes each with a contact angle of $57 \pm 1^\circ$. Membrane hydrophobicity was characterized in terms of sessile drop contact angle measurement by placing a droplet of ultrapure water (5 μ L) onto the membrane surface. The measurement was performed using a VCA2500 XE instrument (AST). Each contact angle was measured three times and an average value was calculated. Prior to the experiment, the membranes were compacted using deionized water until stable permeate flow was achieved.

8.2.3 Fluorescence Analysis

Fluorescence EEMs of the water samples, obtained from the sampling points indicated in Figure 8-1, were acquired using a Varian Cary Eclipse Fluorescence Spectrofluorometer (Palo Alto, CA) collecting 301 individual emission intensity values (within the 300 - 600 nm emission range) at sequential 10 nm increments of excitation wavelengths between 250 nm and 380 nm. Disposable UV-grade polymethylmethacrylate (PMMA) cuvettes with four optical windows were used in the analyses. The PMMA cuvettes, used in this study, gradually filter the emission signals captured below the excitation wavelength (Ex): 285 nm and therefore the fluorescence intensities at emission wavelength (Em) range: 300 – 600 nm captured below Ex: 285 nm were seen to be lower than emission intensities captured using quartz cuvettes at the same conditions (results not shown). This approach provides sufficient spectral information necessary to distinguish different fluorescent elements of the NOM and reduces the risk of cuvette contamination as a source of error (Peiris *et al.*, 2008). The following instrument parameters were maintained during the fluorescence signal acquisition: photomultiplier tube (PMT) voltage = 800 V; scan rate = 600 nm/min and excitation/emission slit width of 10 nm each. These parameter settings were identified in a separate study as the optimum instrument settings for obtaining reproducible fluorescence signals, especially for low NOM concentration levels (Peiris *et al.*, 2009). To eliminate water Raman scattering and to reduce other background noise, fluorescence spectra for Milli-Q (Millipore) water, obtained under the same conditions, were subtracted from all spectra. The temperature of the samples was maintained at room temperature (~ 25°C) during the analyses. Since the pH of all the water samples did not change significantly (pH ~7.3 - 8.4), no pH adjustment was made prior to the fluorescence analysis. A separate study

demonstrated that there is no significant difference in the fluorescence EEM intensities (< 2%) of GRW captured in the above pH range (results not shown). This is in agreement with the previously published data (Spencer *et al.*, 2007). Following this procedure, fluorescence EEMs of 128 samples drawn from 15 different UF experiments and 192 samples drawn from 15 different NF experiments were recorded at different filtration time intervals (i.e. 1, 24, 48 and 96 h). During the course of these experiments and before fluorescence analyses, the Raman scattering peak intensity recorded for Milli-Q water at Ex/Em ~ 348 nm /396 nm was examined to identify any significant fluctuations in the performance of the spectrofluorometer lamp or other hardware. No significant changes in this intensity reading (less than 1%) were observed confirming that there were no significant fluctuations in the performance of the spectrofluorometer during this study.

8.2.4 Fluorescence Data Pre-treatment and PC Analysis

The fluorescence EEM of each sample contained 4214 excitation and emission coordinate points. The fluorescence intensity values corresponding to all 4214 coordinate points (spectral variables) of each EEM were rearranged to generate data rows of intensity values (SI Figure S - 8-1). This procedure generated a 128×4214 data matrix from UF experiments (X_{UF}) and 192×4214 data matrix from NF experiments (X_{NF}). Each row of these data matrices corresponded to each sample and the intensity values of the corresponding EEM were arranged over 4214 columns.

The X_{UF} and X_{NF} data matrices were then separately subjected to PCA. PCA is a well-known technique for data compression and information extraction from a large number of variables.

Essentially, PCA extracts a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and mathematically represented by linear combinations of original variables in the X matrix (X_{UF} or X_{NF} matrix in this case). These new variables, referred to as principal components (PCs), are calculated to account for much of the variance present in the X matrix (Wold *et al.*, 1987; Eriksson *et al.*, 2001) and therefore are able to describe major trends in the original spectral data sets X_{UF} and X_{NF} . PCA decomposes the data matrix X as the sum of the outer product of vectors \mathbf{s}_i and \mathbf{p}_i plus a residual matrix E as presented in Equation 8.1.

$$X = \sum_{i=1}^k s_i \cdot p_i + E \quad (8.1)$$

Where k is the number of samples in the X data set. The \mathbf{s}_i vectors are known as scores (i.e. values) on the PCs (i.e. new variables) extracted by PCA. The \mathbf{p}_i vectors are known as loadings and contain information on how the variables (spectral variables in this case) relate to each other (Wold *et al.*, 1987; Eriksson *et al.*, 2001). The scores (\mathbf{s}_i) generated by PCA can be interpreted as projections of the fluorescence spectral variable to a new space spanned by the PCs (i.e. when the fluorescence spectral variables are transformed to PCs, each spectral variable in the X matrix is projected on to the PC space). The coordinates in this PC space are the scores. The set of scores corresponding to a particular PC can be plotted against another set of scores corresponding to another PC and such plots are called score plots. Generally, the score plot is built on the basis of the first two principal components since these explain most of the variability in the data. The PCs in the PC space are related to the spectral variables in the X matrix (i.e. original variable space) by loadings (Persson and Wedborg,

2001). By examining the loading values related to each PC, it is possible to understand which original spectral variables in the X matrix are better explained by each PC.

Before performing PCA analysis, both X_{UF} and X_{NF} data sets were auto-scaled, i.e. adjusted to zero mean and unit variance by dividing each column by its standard deviation. To determine the number of principal components that are statistically significant in capturing the underlying features in X_{UF} and X_{NF} data sets, a leave-one-out cross-validation method (Eriksson *et al.*, 2001) was implemented. All computations were performed using PLS Toolbox 3.5 (Eigenvector Research, Inc., Manson, WA) within the MATLAB 7.3.0 computational environment (MathWorks, Natick, MA).

8.3 Results and Discussion

8.3.1 Typical Spectral Features in the Fluorescence EEM of GRW

The fluorescence EEM of GRW (i.e. RW) shows a peak (α) at $Ex/Em = 320nm/415nm$ (Figure 8-2), which corresponds to the range reported for fulvic-like HS (Coble *et al.*, 1990; Sierra *et al.*, 2005). The presence of fulvic-like HS in GRW was also independently confirmed by examining the LC-OCD spectra of the same water sample (Peiris *et al.*, 2008). In addition to the primary peak (α), another secondary peak (β) which also corresponds to HS (Sierra *et al.*, 2005; Peiris *et al.*, 2008) appears to be present in the form of a shoulder around $Ex/Em = 270\text{ nm}/460\text{ nm}$ (Figure 8-2). The HS in GRW can be expected to comprise predominantly fulvic acid-type matter compared to humic acid-type matter as reported in other natural waters (Huck, 1999; Sierra *et al.*, 2005). The deviations of the fluorescence

EEM contours seen in the region (Ex/Em: 280nm /330 nm) indicated by δ are believed to be due to the presence of protein-like substances in the water. The existence of a fluorescence EEM peak around the same region (δ) has been previously observed for protein-like substances (Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003). The protein-like peak in the δ region is not clearly visible due to the very low concentration levels of the protein-like substances present in GRW. The light scattering regions (first order Rayleigh scattering region and second order Rayleigh scattering region) observed in the fluorescence EEM are also important areas that provide information related to the particulate/colloidal matter present in water as will be discussed later.

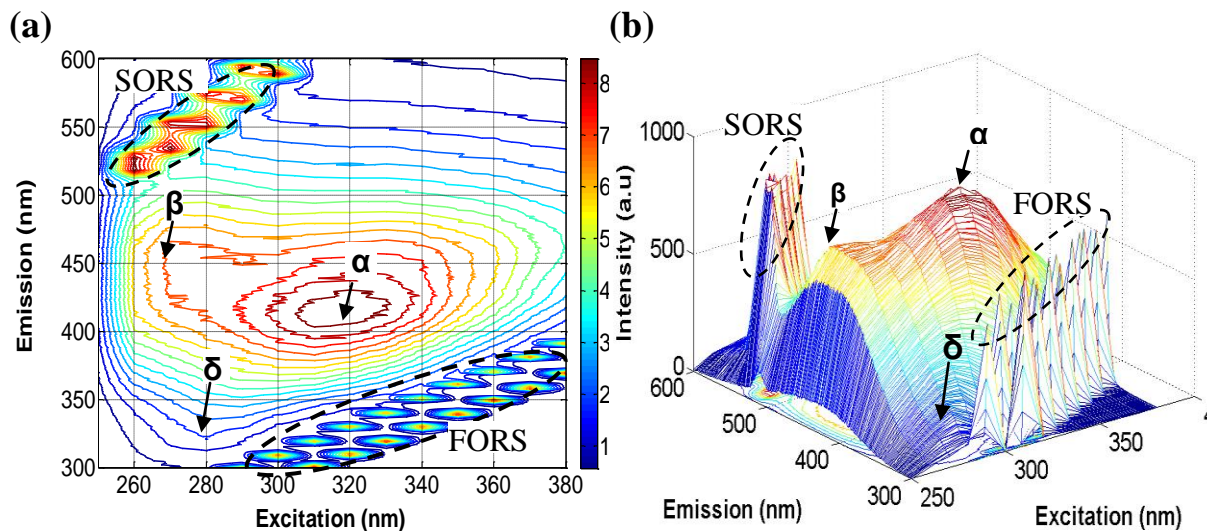


Figure 8-2 Typical fluorescence features seen in the (a) fluorescence EEM for GRW and (b) 3D view of the same EEM. First order Rayleigh scattering (FORS) and Second order Rayleigh scattering (SORS) regions are indicated using dashed-lines.

8.3.2 PCs that Summarize the Total Variance Captured in the Fluorescence EEMs

PCA analyses were performed separately on X_{UF} and X_{NF} matrices to generate new and fewer numbers of variables or PCs to capture any systematic trends present in the 4214

original spectral variables of both X_{UF} and X_{NF} matrices. The first three PCs alone, generated in this way, were able to capture nearly 90% of the total variance present in the original spectral variables of X_{UF} and X_{NF} matrices separately (Table 8-2). The remaining variance (~10%) is due to the combination of unexplained variance by the first three PCs and the instrumental noise in the fluorescence measurements. The instrumental error was however determined to be generally less than 5% for the intensity readings captured by fluorescence EEMs. It is possible to capture this remaining variance by generating more PCs. However, additional PCs were not examined in detail for the reasons explained below (Section 8.3.3).

Table 8-2 Variance captured by the first three principal components.

Principal component	UF spectral data		NF spectral data	
	Variance captured (%)	Cumulative variance captured (%)	Variance captured (%)	Cumulative variance captured (%)
1	66.6	66.6	77.5	77.5
2	12.9	79.5	6.5	84.0
3	8.4	87.9	5.4	89.4

8.3.3 Physical Significance of the PCs Generated by PCA

PCA assigns loading values for each original spectral variable in the X matrix. This process therefore establishes a corresponding loading matrix for each PC. The loading values of each PC denote the relative importance of the fluorescence variables (i.e. excitation-emission wavelength combinations) so that the fluorescence variables with higher intensity values (e.g. fluorescence EEM peaks) of the X matrix are associated with large loading values. Hence, by examining loading matrices, one can understand which original spectral variables in the X matrix, i.e. which combinations of excitation and emission wavelengths, would be most

dominant within the PCs (Persson and Wedborg, 2001). Figure 8-3a, b and c demonstrate the loading values of PC – 1, PC – 2 and PC – 3 that are plotted at their corresponding fluorescence excitation/emission wavelength coordinates. Similar loading plots were generated in the PCA of X_{UF} and X_{NF} but for brevity only the loading plots generated from X_{UF} are demonstrated here.

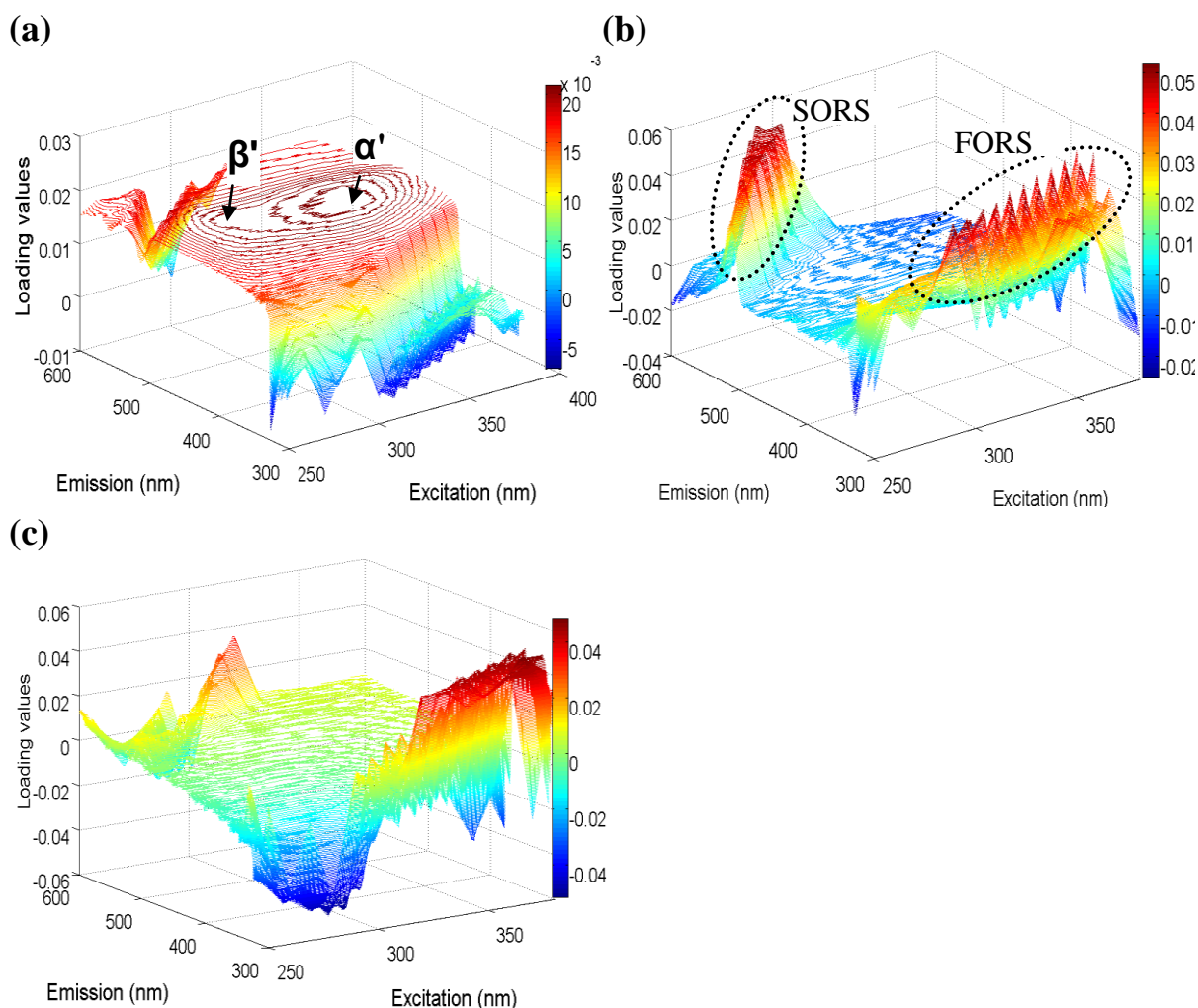


Figure 8-3 Loading plots of (a) PC – 1 - related to the humic content, (b) PC – 2 - related to the particulate/colloidal content and (c) PC – 3 - related to the protein content in water. PCA assigned loading values for each original spectral variable in the X matrix. These loading values are plotted here at their corresponding fluorescence excitation/emission wavelength coordinates. FORS - First order Rayleigh scattering and SORS - Second order Rayleigh scattering regions are indicated using dashed lines.

In the loading plot of PC – 1, a main loading peak (α') at Ex/Em \sim 320 nm/415 nm and second loading peak (β') in the form of a shoulder around Ex/Em = 270 nm/460 nm can be observed (Figure 8-3a). The presence of these loading peaks α' and β' at the same locations where the fluorescence EEM peaks of α and β (Figure 8-2) for HS are situated, therefore indicates that PC – 1 is mostly correlated with the HS content in the water; i.e. samples with high HS content are associated with high PC – 1 scores. The loading plot of PC – 2, on the other hand, demonstrates an array of peaks at the same regions where the light scattering regions (first and second order Rayleigh scattering) are situated in the fluorescence EEM of GRW (Figure 8-2). The intensity values of these light scattering regions increase with increasing particulate/colloidal matter present in the water. Hence, samples with high particulate/colloidal content are associated with high PC – 2 scores. The loading plot of PC – 3 demonstrates a distinct valley at the same regions (Ex/Em: 280 nm/330 nm) where the fluorescence EEM peaks related to protein-like substances occur (Chen *et al.*, 2003; Her *et al.*, 2003). Figure 8-2 also indicates the presence of protein-like substances in GRW in terms of the deviation in the fluorescence EEM contours around the region highlighted by δ . For these reasons, it is reasonable to conclude that PC – 3 is mostly associated with protein-like substances in water. The existence of a valley as opposed to a peak at Ex/Em: 280 nm/330 nm of the loading plot of PC – 3, implies that PC – 3 is inversely related to the protein-like content in the water.

The loading plots of additional PCs were found to contain largely random variation of the loading values. Also, these loading plots did not contain regions that could be related to the spectral regions in the fluorescence EEM of GRW demonstrated in Figure 8-2. For these

reasons, the first three PCs were deemed to be the only PCs that contained sufficiently meaningful information and the rest of this discussion therefore focuses on these.

8.3.4 Performance of the Pre-treatment and UF/NF Stages as Summarized by the Score Plots of PCs and Potential as Performance Monitoring Tool.

The impact of the pre-treatment on the subsequent UF stages was investigated by classifying the filtration operating conditions as normal, when no rapid permeate flux decline was observed (i.e. minimal fouling), and as high fouling events when a decline in flux was observed. Under normal filtration conditions, a total of 28, 28, 12, 16 and 28 samples of RW, RF, B1, B2 and UFp (Figure 8-1), respectively were considered. High fouling events were observed for UF experiments denoted by UF8, UF9, UF12 and UF13. The set of score values for each PC generated by the PCA of the X_{UF} matrix is illustrated in Figure 8-4a (PC – 2 vs. PC – 1) and Figure 8-4b (PC – 3 vs. PC – 2) according to the sample location and specific UF experiment. Note that each value in this score set is directly related to the fluorescence EEM data of each sample in the X_{UF} matrix. The scores corresponding to samples of RW, RF, B1, B2 and UFp formed groups (or clusters) and these groups are indicated by dashed ellipses based on the 95% joint confidence regions (JCRs) of the scores in each group. The calculation of these JCRs, based on the PC scores, was done to define regions related to normal operating conditions of the filtration. The PC scores corresponding to the normal operating conditions of the filtration (28, 28, 12, 16 and 28 samples of RW, RF, B1, B2 and UFp, respectively) were considered in the calculation of these JCRs. The horizontal and vertical orientation of these confidence region ellipses in Figure 8-4a and b is due to the PCA methodology whereby the resulting PC's are orthogonal to each other, i.e. there is

mathematically zero covariance among them. The points denoted by UF8, UF9, UF12 and UF13 indicate high fouling events captured by fluorescence EEMs after 1 h of UF membrane operation.

It should be noted that when only the intensity of the peak maxima of the fluorescence EEMs such as peaks (α), (β), (δ) and the Rayleigh scattering peaks were used in the PCA as opposed to full fluorescence EEMs, the 95% JCRs of the above mentioned groups were not generally separable (i.e. more overlapping regions) unlike the case presented in Figure 8-4a and b (SI Figure S - 8-2). This is expected and explained by the reasoning provided in previous studies (Chen *et al.*, 2003; Stedmon *et al.*, 2003); a smaller number of fluorescence EEM co-ordinates lack the ability to capture the heterogeneity of the different NOM fractions in water. Thus, the use of the full spectra results in better sensitivity in separating the data corresponding to normal operating conditions versus the data measured during fouling conditions.

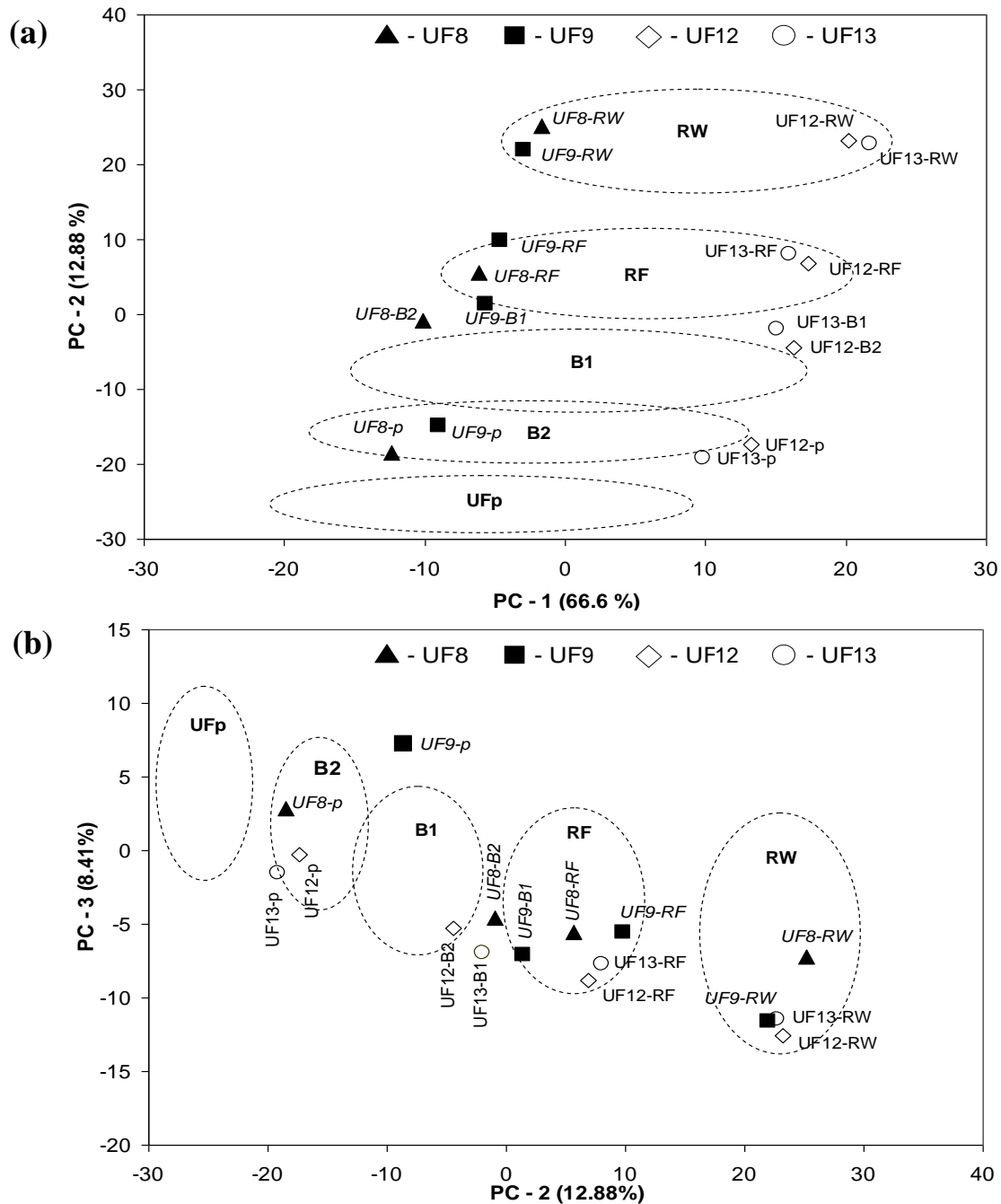


Figure 8-4 Score plots: (a) PC – 2 vs. PC – 1 and (b) PC – 3 vs. PC – 2. PC scores are grouped and named based on the sampling locations: RW, RF, B1, B2 and UFp (Figure 8-1). These groups are indicated by dashed-ellipses showing the 95% joint confidence regions of the scores in each group. UF8, UF9, UF12 and UF13 indicate high fouling events captured within 1 h of UF membrane operation.

The score plot PC – 2 vs. PC – 1 (Figure 8-4a) demonstrates the possibility of defining different regions, which can be considered as normal operating regions, for the roughing filter, two biofilters and the UF step. This information can be further investigated in the context of specific NOM fractions and the corresponding pre-treatment and membrane operation. The 95% JCRs of RW, RF, B1, B2 and UFp demonstrate a progressive shift towards lower values (scores) of PC – 1 and PC – 2. The small shift of PC – 1 indicates limited removal of HS corresponding to a slight shift along the PC – 1 axis while the more pronounced shift along the PC – 2 axis indicates a significant removal of particulate/colloidal matter at each pre-treatment stage and by the UF step. It should be recalled that the treatment steps are sequential (Figure 8-1), except for B1 and B2 that operate in parallel, with B2 having the longer EBCT. The HS removal in these pre-treatment steps, however, cannot be considered as significant due to the overlapping 95 % confidence intervals (CIs) on the PC – 1 axis of Figure 8-4a. The limited HS removal deduced from the shift along the PC – 1 axis is also supported by LC-OCD analyses (Table 8-3), and is consistent with the literature (e.g. Holzalski *et al.*, 1995; Hallé *et al.*, 2009). In particular, the average percentage HS removal by these pre-treatment steps, as calculated from LC-OCD measurement during this study, was less than 10% during this study.

Table 8-3 Comparison of typical TOC, DOC and turbidity values under normal filtration conditions and the values recorded under high fouling events. These values were recorded after one hour of membrane filtration.

		TOC (mg-TOC/L)	DOC (mg-DOC/L)	Turbidity (NTU)
RW	Typical	7.4 - 6.0	7.1 - 5.5	60.2 - 3.2
	UF-8	5.8	5.8	5.7
	UF9	6.6	6.2	2.7
	UF-12	6.6	6.2	15.6
	UF-13	5.6	5.4	7.5
	NF-8	5.8	5.8	5.7
RF	Typical	7.1 - 4.7	6.9 - 4.7	7.8 - 0.7
	UF-8	5.7	5.4	1.2
	UF9	6.3	6.0	1.7
	UF-12	6.5	6.1	3.8
	UF-13	5.9	5.5	3.8
	NF-8	5.7	5.4	1.2
B1	Typical	7.0 - 5.4	7.0 - 5.5	2.1 - 0.1
	UF-8			
	UF9	5.7	5.7	0.9
	UF-12			
	UF-13	5.5	5.3	1.0
B2	Typical	6.7 - 4.1	6.5 - 4.1	1.7 - 0.0
	UF-8	5.4	5.0	0.3
	UF9			
	UF-12	5.7	5.5	2.2
	UF-13			
	NF-8	4.9	4.9	1.2
UFp	Typical	7.7 - 5.2	7.7 - 5.5	0.3 - 0.0
	UF-8	6.0	7.3	0.1
	UF9	5.7	5.8	0.2
	UF-12	5.3	5.4	0.4
	UF-13	5.3	5.3	0.4
NFp	Typical	1.0 - 0.1	0.6 - 0.1	0.4 - 0.1
	UF-8	0.3	0.3	0.4

Typical - denotes the normal filtration conditions; UF-8, UF-9, UF-12, UF-13 and NF-8 are the high fouling events.

The significant removal for particulate/colloidal matter at the pre-treatment stages and by the UF step deduced from the shift along the PC – 2 axis is also supported by the turbidity data presented in Table 8-3. The 95% CIs of RW, RF, B1, B2 and UFp on the PC – 2 axis of Figure 8-4a, are narrow enough to demarcate different operating regions for different treatment stages, with the exception of a small overlapping region between the CIs of B1 and B2, which could be expected. The wider 95% CIs of RW, RF, B1, B2 and UFp, manifested on PC – 1 axis of Figure 8-4a, could be due to the large seasonal variation in the humic content in GRW during the study period; in general, higher humic concentration levels were recorded towards the latter part of January 2008 and lower humic concentration levels were recorded in September 2007.

The score plot PC – 3 vs. PC – 2 (Figure 8-4b) provides information on the reduction of protein-like matter (i.e. higher PC – 3 scores), from RW to UFp, even though the 95% CIs on the PC – 3 axis of Figure 8-4b overlap with each other. There is essentially no removal of protein-like matter by the roughing filter, as would be expected. BF2 shows superior performance to BF1 as indicated by the JCRs of B2 and B1, which is consistent with removal interpretations and fouling data presented by Hallé *et al.* (2009). The removal of protein-like material by the ultrafiltration membrane step is consistent with fouling interpretations provided by Hallé *et al.* (2009) and Haberkamp (2008).

When PCA was performed on the fluorescence EEMs obtained during the NF experiments (i.e. X_{NF} matrix), PC-based 95% JCRs with similar features to those obtained with UF experiments were generated (Figure 8-5a and b). The calculation of these JCRs was based on

the PC scores related to normal operating conditions. The PC scores corresponding to a total of 31, 31, 15, 16, 31, 31 and 31 samples of RW, RF, B1, B2, NF_tank, NF_C and NFp (Figure 8-1) respectively were considered in the calculation of these JCRs. In contrast to the UF experiments, only one high fouling event (NF8) was recorded during NF experiments. The shapes and the coordinates of these JCRs (Figure 8-5a and b) are however dissimilar to those of the JCRs obtained with UF experiments. This dissimilarity is due to differences in the size of PC scales generated by PCA in both cases. As in the PC plots for UF, these JCRs also indicated comparable trends in the HS, particulate/colloidal and protein-like matter removal by different pre-treatment stages located upstream of the NF membrane step. Therefore for brevity, important interpretations of the NF experiments, as explained by PC score plots for NF, are summarized here.

1. A significant level of HS and particulate/colloidal matter removal, as would be expected (Her *et al.* 2007), was seen in the pre-treatment and NF stages.
2. NF was the main contributor to HS removal. Particulate/colloidal matter, on the other hand, was removed to a large extent by the pre-treatment stages. The turbidity data recorded during NF experiments also support this observation (Table 8-3).
3. NF permeate quality was consistent in spite of comparatively larger variation in the membrane feed.
4. Higher rejection levels of protein-like matter were demonstrated by the NF step compared to the pre-treatment stages.
5. The rejection of HS and protein-like matter was generally higher with the TS80 membranes compared to XN45 membranes. The same observation was also made in a related study (Peiris *et al.*, 2008).

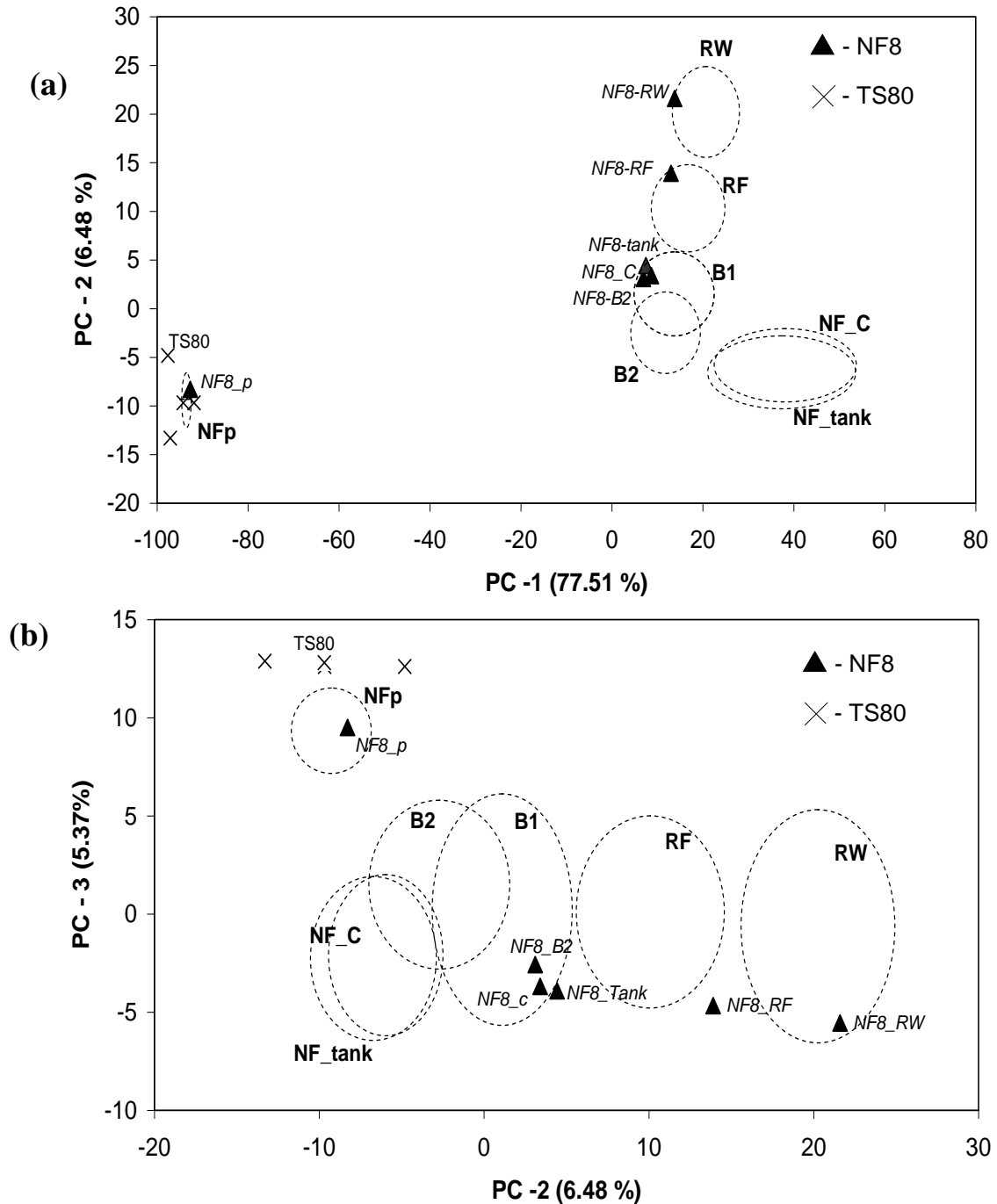


Figure 8-5 Score plot: (a) PC – 2 vs. PC – 1 and (b) PC – 3 vs. PC – 2. PC scores are grouped and named based on the sampling locations: RW, RF, B1, B2, NF_C, NF_tank and NFp. These groups are indicated by dashed-circles/ellipses showing the 95% confidence regions of the scores in each group. NF8 indicates a high fouling event captured within 1 hour of NF membrane operation. Scores of NFp corresponding to the NF experiments run with TS80 are indicated by symbol - “X”.

8.3.5 Identifying High Fouling Events by PCA of Fluorescence EEMs

As indicated earlier, the JCRs in Figure 8-4a and b were generated from the fluorescence EEMs captured during UF experiments for normal operations, i.e. where incidents of high fouling did not occur. Experiments UF8, UF9, UF12 and UF13, on the other hand, exhibited very high fouling within 30, 60, 15 and 10 h of the start of membrane ultrafiltration, respectively. Experiments UF8 and UF9 were performed when HS content was lower (November 2007) as signified by PC – 1 scores towards the lower end of the 95% CI for RW (Figure 8-4a). The HS content in GRW during experiments UF8 and UF9 varied between 3.3 and 3.5 mg C/L, based on LC-OCD determinations. In contrast, experiments UF12 and UF13 were conducted when GRW had a much higher HS content (February 2008) as indicated by the high PC – 1 scores in Figure 8-4a. The HS content in GRW varied between 4.6 and 5.4 mg C/L during experiment UF12, and 4.9 and 5.2 mg C/L during experiment UF13, based on LC-OCD determinations.

During these experiments, HS, particulate/colloidal and protein-like matter in the RF effluent were similar to those within the normal operating conditions of RF as demonstrated in Figure 8-4a and b. This observation is also supported by TOC, DOC and turbidity measurements of RF, recorded during these experiments (Table 8-3). Nevertheless, the effluents of the biofilters in experiments UF8, UF9, UF12 and UF13 had higher PC – 2 scores than the corresponding JCRs of B2 and B1 in Figure 8-4a, indicating reduced particulate/colloidal removal by both biofilters. It is evident that B2 was impacted more than B1. Because of increased levels of particulate/colloidal matter in the UF influent, permeate colloidal levels after the UF stage were also higher as indicated by the much higher PC – 2 scores for UFp,

well outside the normal JCR (Figure 8-4a). The reduced removal of particulate/colloidal matter in the biofilters and UF stage were however not as clearly demonstrated by the turbidity measurements recorded after one hour of UF operation. Turbidity values recorded for B1, B2 and UFp under these high fouling events fell within the general ranges recorded during normal UF conditions (Table 8-3) indicating that turbidity is not a suitable parameter to capture the reduced particulate/colloidal matter removal levels. In addition, the effluent concentrations of protein-like matter from both biofilters, were not very different from their RW values (Figure 8-4b). Therefore the protein-like matter content of the UF permeate also did not differ from the normal operating range. For these reasons, it is reasonable to conclude that the high fouling incidents experienced during experiments UF8, UF9, UF12 and UF13 were due to the poor removal of particulate/colloidal matter during biofiltration pretreatment. This poor performance was linked to the decrease in biofilter activity at low water temperatures (Hallé *et al.*, 2009).

Similar to the high fouling events during UF, poor removal of particulate/colloidal matter seems to also have contributed to the only high fouling event (NF8) recorded during the NF experiments. The biofilter effluent of this experiment fell outside and above the JCR for B2 in the score plot of PC – 2 vs. PC – 1 (Figure 8-5a) and hence indicated a lower than normal level of particulate/colloidal matter removal by BF2 (since the RW level was in the normal range). Turbidity measurements, recorded after one hour of NF operation, did not provide an indication of this poor removal level. Removal of protein-like material, on the other hand, seems to have been normal (Figure 8-5b).

The PC scores (Figure 8-4 and Figure 8-5) therefore clearly indicate a relationship between high fouling events for both UF and NF stages and reduced removals by the biofiltration pretreatment. In particular, as mentioned earlier, it is reduced removals of particulate/colloidal matter that contributed to the high fouling events observed in this investigation. Removal of protein-like material by the biofilters was within the normal operating range. The PC scores that demonstrate these deviations in the performance levels of the biofilters and the subsequent membrane stages were generated by PCA of the fluorescence EEMs obtained just after one hour of filtration. The high fouling event for these membranes however became evident only much later in terms of the increase in the TMP. Therefore it is proposed that PCA of fluorescence EEM data could serve as an early detection method to monitor changes in the membrane feed that could lead to high fouling situations.

8.3.6 Potential for Process Monitoring and Intervention

In this study, PCA of fluorescence EEM was able to capture the differences in UF and NF membrane feedwater that were responsible for changes in fouling rate. An important reason for this capability is the scope of sensitivity: e.g., as discussed above, a higher sensitivity than turbidity measurements for capturing differences in colloidal/particulate matter in the biofilter effluents. Moreover, with the appropriate instrumental parameter settings, it is possible to obtain reproducible fluorescence EEMs even for the NF permeates (Peiris *et al.*, 2008). This means that this approach could be used to monitor membrane permeate with very low NOM concentration levels. In contrast, most other reported NOM characterization techniques require pre-concentration steps prior to the analysis of water with low NOM concentrations, thereby increasing the chances for higher measurement noise.

The fluorescence EEMs obtained during this study were made using off-line measurements, and the signal acquisition time for each EEM was about 5 min. Therefore, as demonstrated above, this approach could be readily used for off-line monitoring of membrane filtration and related pre-treatment processes with relatively inexpensive investments in a spectrofluorometer, computer and related software. It is also possible that fibre optics or robotic sampling could be used to develop an on-line approach. Since the time frames involved with membrane fouling in drinking water treatment applications would normally be expected to be on the order of hours or more, the approach discussed here could be used for near real-time or rapid off-line monitoring. A change in membrane feed water quality leading to accelerated fouling could then be detected in sufficient time that intervention strategies to reduce fouling, such as reducing membrane flux, could be executed.

8.4 Conclusions

This investigation employed PCA of fluorescence EEMs to quantify the impact of pre-treatment stages on the removal of foulants for UF/NF membranes. The following conclusions can be drawn:

1. The performance of biofiltration pre-treatment prior to membrane filtration stage could be monitored, in terms of the removal levels of key membrane foulants such as humic substances, protein-like and particulate/colloidal-like matter by examining the principal components generated by the PCA.

2. The necessary information could be captured by three principal components. Scores on PC – 1 and PC - 2 were largely related to the humic substances level and the particulate/colloidal-like content respectively. Scores of PC – 3 were inversely related to the protein-like content of the water.
3. The approach was able to provide early warning of high membrane fouling events. The fluorescence EEM-based PC score plots, obtained just after 1 h of UF and NF operation, were able to link the high fouling events seen in this study to reduced removals by the biofilters of high influent levels of particulate/colloidal-like material in certain runs. The turbidity measurements made at the same time did not provide an indication of these high fouling events.
4. The approach is very sensitive, as is evident by its ability to be used in analyzing NF permeates containing low levels of organic carbon.
5. This method has the potential to be used as a monitoring tool for membrane-based water treatment and pre-treatment operations, and as an early detection method to identify high fouling events that may arise. This would allow membrane operational changes to be made proactively. In contrast to chromatographic methods, this off-line monitoring approach allows for nearly real-time monitoring.

8.5 Supplementary Information

Table S - 8-1 Details of the ultrafiltration experiments.

Season	#	Type of feed	Permeate flux (LMH)	High fouling events observed after (hours)
Summer 2007	UF5	B2	57	N. A.
	UF6	B2	57	N. A.
	UF7	B1	57	N. A.
Fall 2007	UF8*	B2	57	60 hours
	UF9*	B1	57	30 hours
Winter 2008	UF12*	B2	34	15 hours
	UF13*	B1	40	10 hours
Spring 2008	UF14	B2	46	N. A.
	UF15	B1	47	N. A.
Summer 2008	UF16	B2	57	N. A.
	UF17	B1	57	N. A.
	UF18	B2	57	N. A.

* -denotes high fouling events; N. A. - not applicable

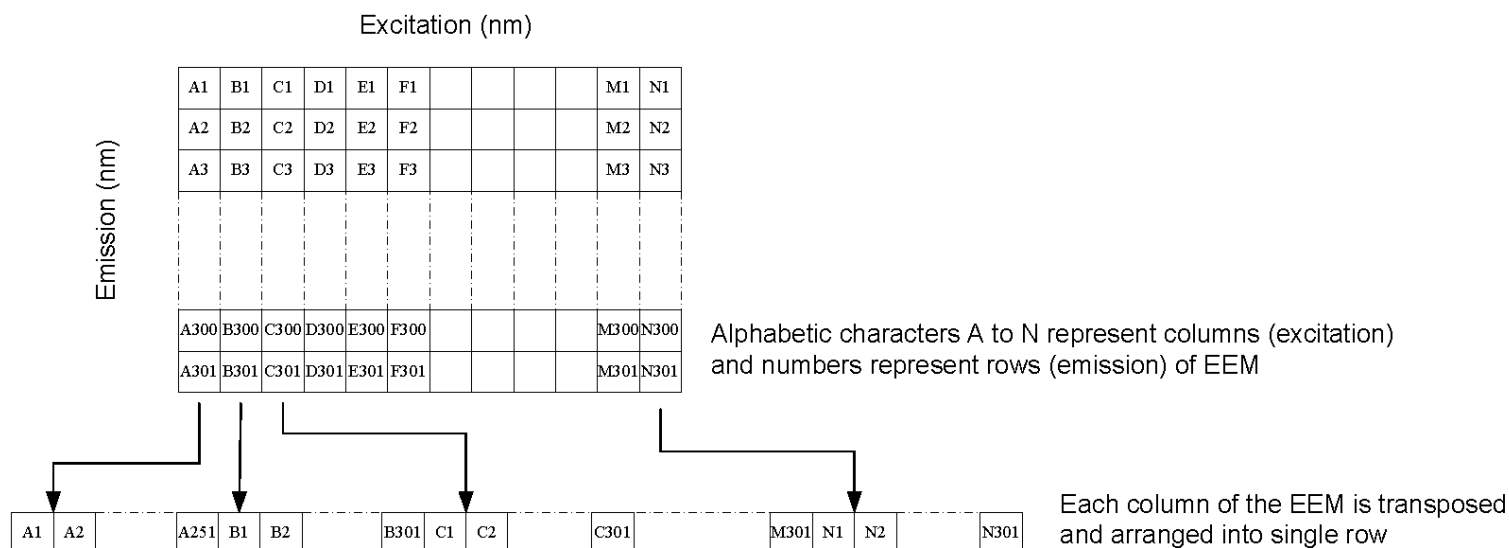
Table S - 8-2 Details of the nanofiltration experiments.

Season	#	Type of feed	Type of membrane	High fouling events observed after (hours)
Summer 2007	NF4	B1	XN45	N. A.
			TS80	N. A.
	NF5	B2	XN45	N. A.
			TS80	N. A.
	NF6	B2	XN45	N. A.
Fall 2007	NF7	B1	XN45	N. A.
	NF8*	B2	XN45	24 hours
	NF9	B1	XN45	N. A.
Winter 2008	NF10	B2	XN45	N.A.
	NF11	B1	XN45	N. A.
	NF12	B2	XN45	N. A.
	NF13	B1	XN45	N. A.
Spring 2008	NF14	B2	XN45	N. A.
	NF15	B1	XN45	N. A.
Summer 2008	NF16a	B2	XN45	N. A.
	NF16b	B2	TS80	N. A.

* -denotes high fouling events; N.A. - not applicable

Step-1

Rearrangement of fluorescence EEM intensity values to a row



Step-2

The same procedure is followed for every sample to develop a predictor data set (X)

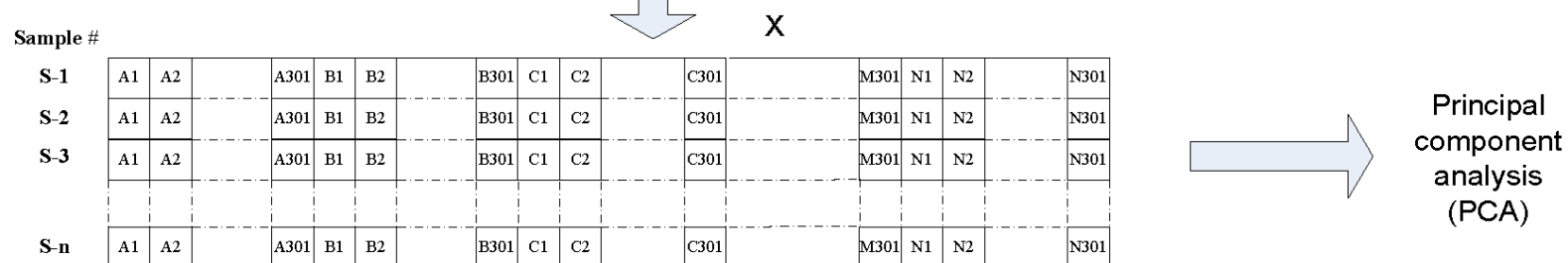


Figure S - 8-1 Steps involved in generating data rows of intensity values for principal component analysis (PCA).

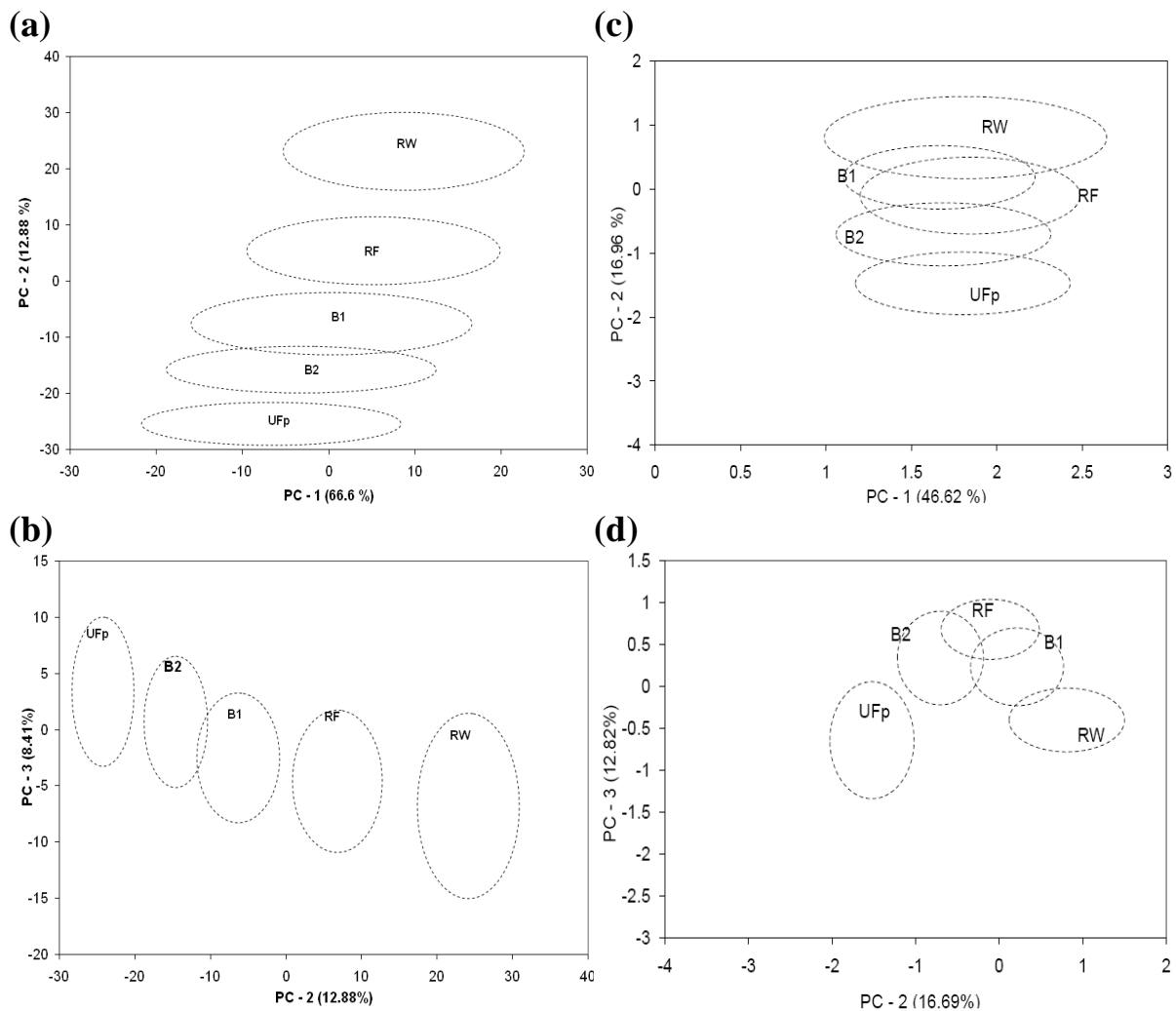


Figure S - 8-2 A comparison of the JCR obtained by (a,b) PCA of the full fluorescence EEM intensity data and (c,d) PCA of only the main peak positions of the fluorescence EEMs such as peaks (α), (β), (δ) and the Rayleigh scattering peaks recorded during the UF experiment.

CHAPTER 9

Understanding Fouling Behaviour of Ultrafiltration Membrane Processes and Natural Water using Principal Component Analysis of Fluorescence Excitation-Emission Matrices*

Overview

Membrane fouling is a major obstacle for maintaining efficient membrane-based drinking water treatment processes. Natural organic matter (NOM) foulant fractions such as humic substances (HS) and protein-like matter as well as colloidal/particulate matter are known to be the major membrane foulants in the ultrafiltration-based drinking water processes. In this study, a fluorescence excitation-emission matrix (EEM) approach was used for characterization of these major foulant fractions. Unlike most NOM and colloidal/particulate matter characterization techniques, this method can provide fast and consistent analyses with high instrumental sensitivity. Principal component analysis (PCA) was used to extract principal components (PCs) that contained information relevant for membrane fouling from fluorescence EEM measurements collected during a bench-scale cross-flow ultrafiltration process. These PCs were related to the major membrane foulant groups such as HS, protein-like and colloidal/particulate matter in natural water used in this study. PC score-based

* Peiris, R. H., Budman, H., Moresoli, C., Legge, R. L. This chapter has been accepted for publication in the Journal of Membrane Science (Accepted date: March 31, 2010; DOI: 10.1016/j.membsci.2010.03.047).

analysis revealed that colloidal/particulate matter mostly contributed to reversible fouling. HS and protein-like matter were largely responsible for irreversible fouling behaviour. The proposed method proved suitable for identifying the major foulant fractions and their contribution to reversible and irreversible membrane fouling, illustrating its potential for monitoring and controlling membrane fouling in drinking water treatment applications.

Keywords: Membrane fouling, Fluorescence spectroscopy, Principal component analysis, Natural water, Ultrafiltration

9.1 Introduction

Application of ultrafiltration (UF) membrane processes are increasing for the production of drinking water. Incorporation of this technology allows effective removal of particulate matter as well as a smaller footprint for the treatment facility. UF of natural water can also reduce natural organic matter (NOM) and pathogenic organisms in water when combined with other pre- and post-treatment methods like coagulation and disinfection for drinking water production (Fiksdal and Leiknes, 2006). In addition, UF membrane systems can serve as a pre-treatment step for nanofiltration (NF) and reverse osmosis (RO) membrane systems to remove coarser particulate/colloidal material that could contribute to fouling of NF and RO membranes.

Fouling of UF membranes however is a major constraint for both implementation and efficient performance in the production of drinking water. NOM and colloidal/particulate matter are considered to be the major membrane foulants (Saravia *et al.*, 2006; Jermann *et*

al., 2007) in the UF of surface and ground water for the production of drinking water. NOM in water contains a wide range of structurally complex compounds that result from prolonged degradation of plant and animal materials (Thurman, 1985) NOM is also considered to be the main component of organic carbon in aquatic systems (Auostin *et al.*, 2001). The major fraction of NOM, which comprises over 50% of the dissolved organic carbon (DOC) content, is composed of humic substances (HS)-like matter (Thurman, 1985). Protein, amino acids, transphilic acids and polysaccharides fall into the non-humic part of NOM which account for 20 - 40% of the DOC content in natural waters (Fan *et al.*, 2001). Fouling is generally caused by the slow accumulation of colloidal/particulate matter and NOM on the surface and in the pores of the membranes over time. This contributes to increased operational costs as a result of permeate flux decline and/or increased trans-membrane pressure (TMP) requirements. In addition, the frequent chemical cleaning associated with the long term use of membranes for the production of drinking water leads to the deterioration of membrane performance, shortened membrane service life and increased costs.

9.1.1 Natural Water and Membrane Foulants

The major NOM fractions, such as HS, protein- and polysaccharide-like substances as well as colloidal/particulate matter present in natural water contribute to hydraulically reversible and irreversible membrane fouling (Jermann *et al.*, 2008a and 2008b). These major NOM fractions occur largely in dissolved form (Thurman, 1985) while a minor portion of protein- and polysaccharide-like matter is found in colloidal/particulate form. (Amy, 2008). Colloidal/particulate matter in natural water is comprised of NOM-like material and inorganic components (Auostin *et al.*, 2001; Jermann *et al.*, 2008b). HS are widely considered

to be the most detrimental NOM-foulants, causing irreversible fouling by membrane adsorption and pore blocking for UF membranes (Jucker and Clark, 1994; Jones and O'Melia, 2000; Auostin *et al.*, 2001; Jermann 2007 and 2008b). Protein- and polysaccharide-like matter have also been observed to contribute to irreversible UF membrane fouling (Jones and O'Melia, 2000; Auostin *et al.*, 2001; Jermann *et al.*, 2007). Although some studies have suggested that protein- and polysaccharide-like material have less fouling potential than HS (Jones and O'Melia 2000; Jermann *et al.*, 2007), their preferential fouling in comparison to HS will depend on the hydrophilicity of the membranes (Kaiya *et al.*, 2000; Her *et al.*, 2007). The individual contribution of colloidal/particulate matter, on the other hand, has been identified to be largely responsible for reversible fouling although the interaction between colloidal/particulate matter and NOM could also result in irreversible fouling behaviour in UF membranes (Jermann *et al.*, 2008b). Characterization of these membrane foulant components is therefore essential for understanding membrane fouling and development of fouling monitoring and control strategies (Amy, 2008).

9.1.2 Fluorescence-based Fouling Monitoring Approach

The fluorescence excitation emission matrix (EEM) approach is being increasingly used in NOM characterization (Hudson *et al.*, 2007; Henderson *et al.*, 2009). The fluorescence EEM method generated a large number of fluorescence intensity readings at different combinations of excitation and emission wavelengths. Compared to single excitation/emission-scan or synchronous fluorescence spectroscopic approaches, the fluorescence EEM method provides more detailed information (Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003; Sierra *et al.*, 2005) and better sensitivity (Peiris *et al.*, 2010) for capturing detailed and subtle fluorescence

features that correspond to HS and protein-like materials. The light scattering regions that can be collected in the fluorescence EEMs can also provide information related to the particulate/colloidal matter present in water (Peiris *et al.*, 2010). In addition, compared to other available NOM characterization methods (Huber *et al.*, 1992; Her *et al.*, 2003; Gray *et al.*, 2007), this approach is suitable for differentiating the major NOM fractions and is suitable for rapid, direct and accurate analysis with high instrumental sensitivity (Coble *et al.*, 1990; Peiris *et al.*, 2008).

Previous studies have demonstrated that fluorescence EEMs can be used to understand membrane fouling behaviour in drinking water treatment processes (Peiris *et al.*, 2008 and 2010). By analyzing all the spectral information contained in the fluorescence EEMs in terms of the entire array of wavelength-dependent fluorescence intensity readings, it is possible to capture the subtle changes in fluorescence EEMs that are related to changes in the heterogeneity of the different foulant fractions in water. This way of analyzing the fluorescence EEMs represents a more accurate characterization of NOM foulants present in water as opposed to the more common method which examines only the major peaks, i.e. a limited number of excitation-emission coordinate pairs (Persson and Wedborg, 2001; Chen *et al.*, 2003; Stedmon *et al.*, 2003; Boehme *et al.*, 2004).

In our previous work, principal component analysis (PCA) was successfully applied to deconvolute spectral information contained within full fluorescence EEMs into principal components (PCs) that were related to HS-like, protein-like and colloidal/particulate matter present in natural water (Peiris *et al.*, 2010). This approach, which was based on the PC

scores generated by the PCA of fluorescence EEMs, is believed to be suitable for rapid monitoring of the subtle changes in the major membrane foulants and the performance of membrane operations for drinking water applications.

In this study, a fluorescence EEM approach together with PCA was used for characterization of major membrane foulants present in natural water and during membrane cross-flow UF operation. This PCA approach was developed to identify different fouling behaviour resulting from the changes of the major membrane foulants during UF operation with successive permeation and membrane backwashing cycles. Fouling behaviour consisted of reversible fouling (loosely attached) and irreversible fouling (tightly bound to the membrane). Natural river water that is comprised of a complex matrix of different membrane foulant components was used in this study. Three major membrane foulants, HS-like, protein-like and colloidal/particulate fractions were characterized over the course of the UF operation and for membrane feeds of different fouling strengths. In particular, PCs were associated to specific membrane foulant fractions and provided a qualitative understanding of their individual contribution to reversible fouling and irreversible fouling observed for two UF membranes with different rejection characteristics. These findings were corroborated with fluorescence analysis of the foulants extracted from the membranes. Thus, the proposed approach represents a promising avenue for fouling monitoring, control and optimization of membrane systems for the production of drinking water.

9.2 Materials and Methods

9.2.1 Feed Water and Pre-treatment

Grand River water (GRW) (Southwestern Ontario, Canada) was filtered using a 200 micron filter (038A-2080; Keller Products, Inc. Acton, MA) and used as the feed in UF experiments. Typical UF feed water quality parameters recorded during the experimental period are presented in Table 9-1. The water collected from the Grand River was stored at 4 °C before the experiments and always used within 48 hours of the collection time.

Table 9-1 UF feed water quality parameters from March – October, 2009

Parameters	Membrane feed (filtered Grand River water)
Dissolved organic carbon (DOC, mg/L)	3.9 – 6.5
Turbidity (NTU)	1.2 – 3.8
pH	7.8 – 8.4

9.2.2 Bench-scale Membrane Ultrafiltration Set-up

UF experiments were performed with a bench-scale flat sheet cross-flow set-up as shown in Figure 9-1. The membrane cross-flow cell holder (Sterlitech CF042; Sterlitech Corporation, Kent, WA) had an effective membrane area of 42 cm². The channel width, length and height of the membrane cell holder were 0.0457 m, 0.0920 m and 0.0023 m, respectively. Flat sheet UF membranes with a molecular weight cut-off (MWCO) of 20 kDa (Polysulfone - YMEWSP3001; GE Osmonics) and 60 kDa (Polyethersulfone - YMPWSP3001; GE Osmonics) from Sterlitech Corp. were used in this experimental set-up. The contact angle

measurements of the virgin 20 kDa and 60 kDa membranes were $72 \pm 2^\circ$ ($n = 6$) and $80 \pm 1^\circ$ ($n = 6$), while the initial pure water flux at TMP = 15 psi were ~ 1.6 and ~ 2.4 L/min.m², respectively. It should be noted that the 60 kDa membranes were immersed in 50 % vol/vol ethanol solution (100 ml) for 24 hours followed by rinsing with DI and Milli-Q water (resistivity = 18.2 MΩ cm) before filtration as advised by the supplier. A new membrane was used for each filtration run. Prior to each experiment, the membranes were compacted at 15 psi (103.4 kPa) using Milli-Q water until a stable permeate flow was achieved.

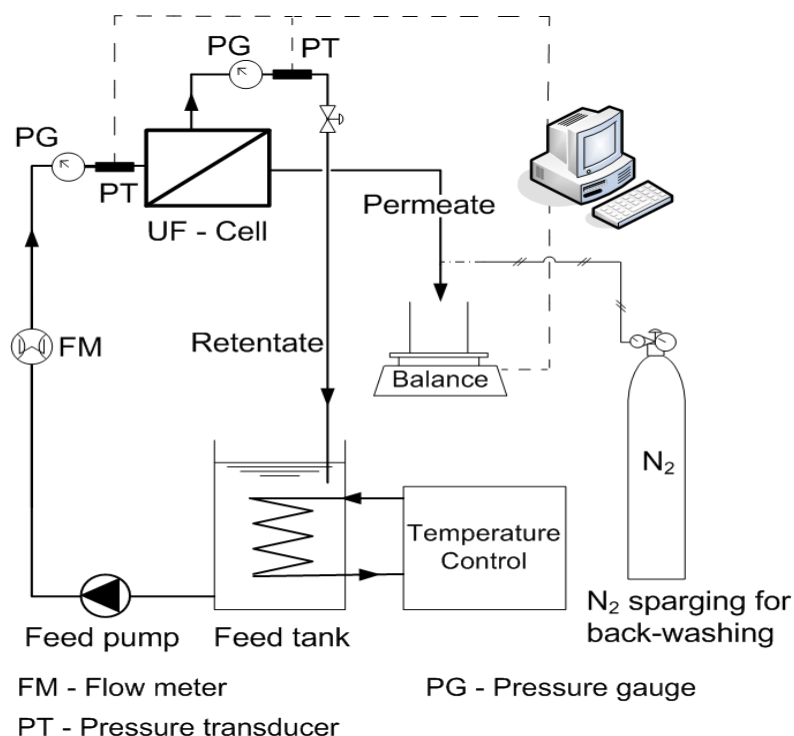


Figure 9-1 Schematic of the bench-scale ultrafiltration cross flow set-up.

Filtered GRW was fed to the membrane set-up at 0.6 L/min and the retentate was circulated back to the feed tank while the permeate was continuously removed. The wall shear stress of the membrane cell holder was ~ 41.3 s⁻¹. The TMP was maintained at 15 psi (103.4 kPa) and the temperature of the feed tank was maintained at $\sim 25 \pm 1^\circ$ C using a temperature controller.

The permeate water mass across the UF membrane was recorded with a balance connected to a computer using a LabView-based (LabView 8.0; National Instruments, Austin, TX) interface. Permeate flux was calculated from the permeate mass measurements. The filtration consisted of a two step operation cycle: (1) permeation period of 1 h and (2) back-washing for 20 s. Back-washing of the membrane was implemented by forcing the permeate contained in the permeate pipe and permeate channels of the membrane cell holder in the opposite direction of the filtration through the membrane using pressurized Nitrogen gas at 10 psi (68.9 kPa). The fluorescence EEMs of both retentate and permeate were recorded at 15 min intervals during the course of the permeation step.

9.2.3 Fluorescence Analysis

The fluorescence EEMs were recorded, using a Varian Cary Eclipse Fluorescence Spectrofluorometer (Palo Alto, CA) by scanning 301 individual emission spectra (300 – 600 nm) at sequential 10 nm increments of excitation wavelengths between 250 and 380 nm. A detailed description of the fluorescence EEM analysis procedure and the selection of the spectrofluorometer parameter settings can be found in Peiris *et al.* (2008 and 2009). Corrections for inner filter effects were not applied as inner filtering effects were not expected to be significant at the low concentration levels (Hudson *et al.*, 2008) that are typical to the GRW used in this study. To eliminate water Raman scattering and to reduce other background noise, fluorescence spectra for Milli-Q water, obtained under the same conditions, were subtracted from all spectra. During the course of the fluorescence analyses, differences in Raman scattering peak intensities recorded for Milli-Q water at Ex/Em ~ 348 nm/396 nm was less than 2%, confirming that there were no significant fluctuations in the

performance of the spectrofluorometer lamp or other hardware. The temperature of the water samples was maintained at room temperature ($\sim 25^{\circ}\text{C}$) during the analysis. The pH of all the water samples did not change significantly ($\text{pH} \sim 7.8 - 8.4$) during the study and no pH adjustment was made prior to the fluorescence analysis. A separate study demonstrated that there was no significant difference in the fluorescence EEM intensities ($< 2\%$) of GRW captured in the above pH range (results not shown). This is in agreement with previous studies (Spencer *et al.*, 2007).

9.2.4 Other Analytical Methods

The DOC content of water was determined using an OI-Analytical TOC analyser (Model 1010, College Station, TX) with a wet-oxidation method as described in Standard Methods (2005) 5310D. Turbidity measurements were performed using a turbidity meter (Hach 2100P) following the Standard Method (2005) 2130. Membrane contact angle measurements were obtained with the sessile drop contact angle measurement method by placing a droplet of Milli-Q water ($5\ \mu\text{L}$) onto the membrane surface and using a VCA2500 XE instrument (AST Products Inc., Billerica, MA).

9.2.5 Extraction of Membrane Foulants

At the end of each filtration, the foulant material considered to be loosely attached on the membrane was gently scrapped off from the membrane by using the tip of a laboratory spatula and dissolved thoroughly in 20 ml of Milli-Q water using a vortex mixer. This was followed by rinsing the membrane surface under running DI water for ~ 2 min to remove any remaining loosely attached foulant material. Subsequently, the membrane was placed on the

membrane cell holder and back-washed for 20 s using Milli-Q water as the back-washing medium. This was achieved by operating the UF system with Milli-Q water as the feed until the permeate pipe and permeate channels were filled with Milli-Q water and then implementing the back-washing procedure as explained in Section 9.2.2. Each membrane was then immersed in 50 ml of ethanol solution (50 % vol/vol) for 24 h to extract the remaining foulant components from the membrane which were not removed by the above cleaning procedure. The fluorescence EEMs of the two foulant extracts (i.e. loosely attached foulant layer dissolved in Milli-Q water and the ethanol-based foulant extract) were obtained following the procedure explained in Section 9.2.3. The fluorescence EEM of the ethanol solution (50 % vol/vol), was subtracted from all fluorescence EEMs of ethanol-based foulant extracts. The pH of these foulant extracts ranged from 7.1 to 8.3 and therefore no pH adjustment was made prior to the fluorescence analysis for the reasons mentioned in Section 9.2.3.

9.2.6 Fluorescence Data Pre-treatment and PCA

The fluorescence EEM of each sample contained 4214 excitation and emission coordinate points. The fluorescence intensity values corresponding to all 4214 coordinate points (spectral variables) of a given EEM were rearranged following the fluorescence EEM data rearrangement procedure described in Peiris *et al.* (2010). This resulted in a $n \times 4214$ fluorescence data matrix, with each row containing fluorescence EEM data points of a given sample; n represents the number of samples corresponding to the combination of retentate and permeate samples obtained at 15 min intervals during the UF permeation experiments as described above. This procedure generated one data matrix for each membrane type referred

heretofore as matrix X60 and matrix X20 for UF experiments performed with 60 kDa and 20 kDa membranes, respectively. The X60 and X20 data matrices contained 525 and 560 fluorescence EEMs from 15 and 16 different UF experiments, respectively. Each UF experiment contained one feed water sample obtained at the beginning of the experiment and 17 samples each of retentate and permeate obtained during the permeation experiments that were analyzed to obtain their fluorescence EEMs. PCA was then performed on matrices X60 and X20 separately to generate PC scores as explained in Persson and Wedborg, (2001) and Peiris *et al.* (2010).

PCA extracts a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and mathematically represented by linear combinations of original variables in the X matrix of a given membrane type (X60 or X20 matrix in this case). These new variables, referred to as PCs, were calculated to account for much of the variance present in the X matrix (Wold *et al.*, 1987; Eriksson *et al.*, 2001) and therefore describe major trends in the original spectral data sets of X60 and X20. PCA decomposes the data matrix X as the sum of the outer product of vectors \mathbf{s}_i and \mathbf{p}_i plus a residual matrix E as presented in Equation 9.1.

$$X = \sum_{i=1}^n s_i \cdot p_i + E \quad (9.1)$$

The \mathbf{s}_i vectors are known as scores (i.e. values) on the PCs (i.e. new variables) extracted by PCA. The \mathbf{p}_i vectors are known as loadings and contain information on how the variables (fluorescence variables in this case) are correlated. The scores (\mathbf{s}_i) generated by PCA can be

interpreted as projections of the fluorescence spectral variables to a new space spanned by the PCs (i.e. when the fluorescence spectral variables are transformed to PCs, each spectral variable in the X matrix is projected on to the PC space). The PCs in the PC space are related to the spectral variables in the X matrix (i.e. original variable space) by loadings. By examining the loading values related to each PC, it is possible to identify which original spectral variables in the X matrix are explained preferentially by a given PC. For a detailed description of PCA see Wold *et al.* (1987) and Eriksson *et al.* (2001).

Before performing PCA analysis, both X60 and X20 matrices were auto-scaled, i.e. adjusted to a zero mean and unit variance by dividing each column by its standard deviation, to assign equal prominence to each of the fluorescence intensity variables in X60 and X20 matrices. To determine the number of PCs that were statistically significant in capturing the underlying features in X60 and X20 data sets, the leave-one-out cross-validation method (Eriksson *et al.*, 2001) was implemented. All computations were performed using the PLS Toolbox 5.2 (Eigenvector Research, Inc., Manson, WA) within the MATLAB 7.8.0 computational environment (MathWorks, Natick, MA).

9.2.7 PC Score-based Evaluation of Membrane Fouling

The statistically significant PCs extracted from the fluorescence data captured during UF experiments, were found to be correlated to the major foulant fractions, HS-like, protein-like and particulate/colloidal matter present in water as will be explained subsequently. The evolution of the PC scores corresponding to these membrane foulant fractions during the UF of water is related to the membrane fouling behaviour (Peiris *et al.*, 2010). Therefore, the PC

scores associated with the retentate ($s_{j,R}$) and permeate ($s_{j,P}$) of the UF process, where s_j represents the PC score related to the j^{th} membrane foulant present in water (discussed later), were used to understand the fouling associated with the 60 kDa and 20 kDa membranes. The difference between the $s_{j,R}$ and $s_{j,P}$ score values (Δs_j) was assumed to be proportional to the accumulation of specific foulant components – related to the j^{th} PC – on the surface and/or in the pores of the membrane (Equation 9.2) of a given time interval (Δt) during the UF permeation experiment. Since the retentate was circulated back to the feed tank (Figure 9-1), Δs_j was normalized with respect to the PC scores of the initial feed water (generated from the fluorescence EEM data of feed water obtained at the start of the experiments) to account for the concentration effect caused by the recirculation of the retentate as indicated in Equation 9.2.

$$\text{Accumulation of foulants} \propto \Delta s_{j, \text{Norm}} = (s_{j,R} - s_{j,P}) \cdot \frac{s_{j, F_o}}{s_{j,R}} \quad (9.2)$$

for $j = 1, 2, 3, \dots, N$

Where $\Delta s_{j, \text{Norm}}$ is the normalized difference between $s_{j,R}$ and $s_{j,P}$ with respect to the PC scores of the initial feed water (s_{j, F_o}) related to the j^{th} PC. Subscripts F_o , R and P denote the initial feed water, retentate and the permeate. N is the number of PCs generated by PCA which are statistically significant and deemed to be important for capturing the information related to the major foulants groups as explained in Section 9.2.6. The evolution of $\Delta s_{j, \text{Norm}}$ with time was examined to understand the different fouling behaviour of UF membranes as explained in a following section.

9.3 Results and Discussion

9.3.1 Typical Fluorescence EEM Features of GRW

The fluorescence EEM of GRW contained three regions or peaks representative of major NOM fractions. There was a typical peak (α) at excitation wavelength (Ex) \sim 320 nm and emission wavelength (Em) \sim 415 nm (Figure 9-2), which corresponds to the range reported for HS-like NOM (Coble *et al.*, 1990; Sierra *et al.*, 2005) and confirmed by previous independent liquid chromatography-organic carbon detection (LC-OCD) analysis of GRW (Peiris *et al.*, 2008 and 2010). A secondary peak (β), which also corresponds to HS (Sierra *et al.*, 2005; Peiris *et al.*, 2008) appears as a “shoulder” to the primary peak (α) at Ex/Em \sim 270 nm/460 nm (Figure 9-2). The deviations of the fluorescence EEM contours seen in the region (δ) around Ex/Em \sim 280 nm/330 nm is indicative of the presence of protein-like substances in the water. The existence of a fluorescence EEM peak around the same region has also been previously reported for protein-like substances in natural water sources (Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003). The protein-like peak in the δ region is less clearly visible and would reflect the very low concentration levels of protein-like substances present in GRW compared to HS. Rayleigh light scattering regions (RS) observed in the fluorescence EEM have been shown to contain information related to colloidal/particulate matter present in water (Peiris *et al.*, 2010).

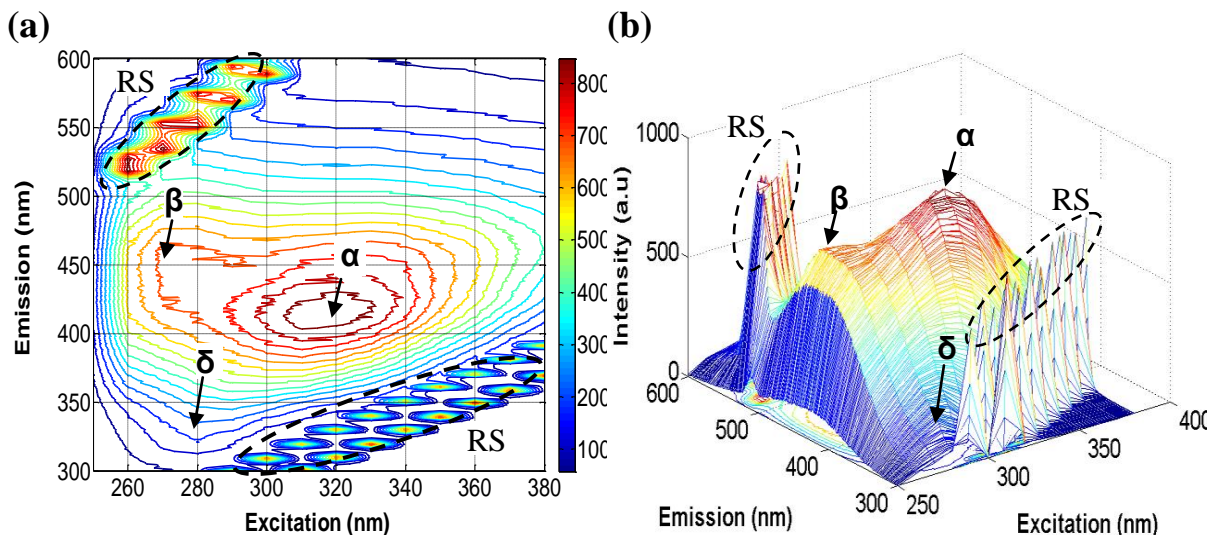


Figure 9-2 Typical fluorescence features seen in the (a) fluorescence EEM for GRW and (b) 3D view of the same EEM. Rayleigh light scattering (RS) regions are enclosed by the dashed lines.

9.3.2 PCA of Fluorescence EEMs

PCA was performed separately on X60 and X20 matrices (i.e. fluorescence data obtained for the 60 kDa and the 20 kDa membrane, respectively) to generate new and fewer numbers of variables or PCs to capture any systematic trends present in the 4214 original spectral variables specific to each membrane type. This process resulted in four statistically significant PCs for the X60 matrix capturing nearly 90% of the total variance present in the raw spectral variables (Table 9-2). In contrast, the X20 matrix contained only three statistically significant PCs that captured approximately 90% of the total variance. The remaining variance ($\sim 10\%$) was due to the combination of unexplained variance by these PCs and the instrument noise (determined to be less than 5% of the intensity readings) in the fluorescence measurements. In principle the remaining variance could be captured by generating additional PCs; however, none of the additional PCs were statistically significant

(< 2% variance captured by each additional PC) and they were not found to be related to membrane foulant fractions present in water as will be discussed later.

Table 9-2 Variance captured in the PCA of X60 and X20 matrices.

Principal component	X60 matrix (60 kDa UF spectral data)		X20 matrix (20 kDa UF spectral data)	
	Variance captured (%)	Cumulative variance captured (%)	Variance captured (%)	Cumulative variance captured (%)
1	63.0	63.0	75.0	75.0
2	16.4	79.4	9.6	84.6
3	5.5	84.9	6.1	90.7
4	4.7	89.6	-	-

9.3.3 PCA of Fluorescence Data Identifies Foulant Components

PCA assigns loading values for each individual spectral variable in the X matrix (i.e. X60 or X20). This process establishes a corresponding loading matrix for each PC so that higher fluorescence intensity values (e.g. fluorescence EEM peaks) of the X matrix are associated with large loading values. Therefore, loading matrices related to each PC could be examined to understand which individual spectral variables in the X matrix, i.e. which excitation and emission wavelength combinations would predominate for a given PC (Persson and Wedborg, 2001). Figure 9-3a, b, c and d illustrate the loading plots of PC – 1, PC – 2, PC – 3 and PC – 4 generated by the PCA of X60 matrix. The loading values in the matrices were plotted at their corresponding fluorescence excitation/emission wavelength coordinates in 3D format. A similar approach was applied to the X20 matrix, where three loading plots

corresponding to the three significant PCs (i.e. PC – 1, PC – 2 and PC – 3) were generated. The loading plots of PC – 1, PC – 2 and PC – 3 generated for the two membrane types had similar features but different magnitudes in loading values (data not shown for X20 matrix).

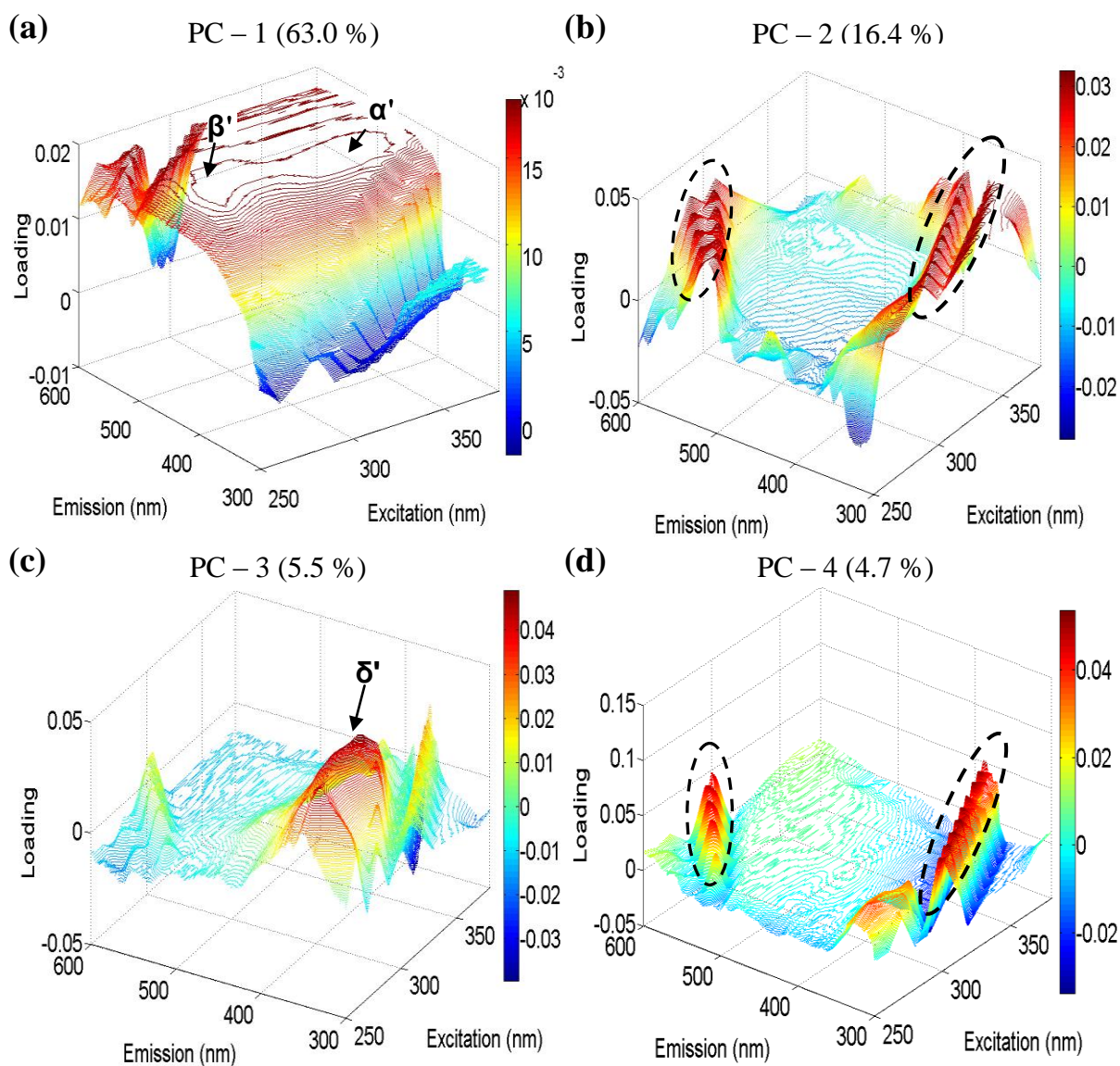


Figure 9-3 3D illustrations of the loading matrices of (a) PC – 1, (b) PC – 2, (c) PC – 3 and (d) PC – 4 generated by the PCA of 60 kDa data. Rayleigh scattering peak-like regions are enclosed by the dashed lines. Variance captured by each PC is indicated within parentheses.

The loading plot of PC – 1 contains a major peak (α') at Ex/Em \sim 320 nm/415 nm and a minor peak (β') in the form of a shoulder around Ex/Em \sim 270 nm/460 nm (Figure 9-3a). The presence of these loading peaks α' and β' at similar locations as the fluorescence EEM peaks of α and β (Figure 9-2) for HS indicate that PC – 1 is mostly correlated with the HS content in the water; i.e. samples with high HS content are associated with high PC – 1 scores. The loading plot of PC – 2 (Figure 9-3b), on the other hand, demonstrates an array of peaks at locations similar to the fluorescence EEM of the RS regions for GRW (Figure 9-2). The intensities of these RS regions increase with increasing colloidal/ particulate content in the water. Hence, samples with high colloidal/particulate content are associated with high PC – 2 scores. The loading plot of PC – 3 (Figure 9-3c) also demonstrates a distinct peak (δ') in the same regions (Ex/Em: 280/330 nm) where the presence of protein-like NOM are indicated in the fluorescence EEM of GRW in Figure 9-2. Therefore it is reasonable to conclude that high PC – 3 scores are associated with high protein-like content in water. In addition, the loading plot of PC – 4 also contains an array of peaks similar in appearance and location to RS peaks captured in the fluorescence EEMs of GRW. This indicates that in addition to PC – 2, PC – 4 also captures meaningful spectral information related to colloidal/particulate matter present in water. Thus, high PC – 4 scores are correlated with high colloidal/particulate content.

The loading plots of additional PCs for 20 kDa and 60 kDa UF experiments (i.e. in addition to the 3 and 4 PCs for 20 kDa and 60 kDa UF experiments, respectively) contained largely random variation of the loading values. These loading plots were also not related to the representative fluorescence EEM regions of the major fractions of GRW (Figure 9-2). The

remaining discussion therefore focuses only on the PCs which were found to be statistically significant and were able to capture meaningful spectral information.

These PC scores were then used to calculate $\Delta S_{j, \text{Norm}}$ according to Equation 9.2 for each PC. Hence, $\Delta S_{1, \text{Norm}}$, $\Delta S_{2, \text{Norm}}$ and $\Delta S_{3, \text{Norm}}$ describe the differences between the retentate and permeate PC scores and are representative of the accumulation of HS-like, colloidal/particulate and protein-like foulant fractions, respectively, on the surface and/or in the pores of the membrane during a given time interval of the permeation for both 60 kDa and 20 kDa UF experiments. In addition to $\Delta S_{2, \text{Norm}}$, $\Delta S_{4, \text{Norm}}$ is correlated to the accumulation of colloidal/particulate substances for the 60 kDa UF operation.

9.3.4 Different Fouling Behaviour Captured by PC Scores

UF experiments were performed with filtered GRW of varying qualities measured in terms of DOC and turbidity analyses and collected between March 2009 and October 2009. Figure 9-4a and b illustrate high, medium and low fouling behaviour experienced by both 60 kDa and 20 kDa membranes due to the changes in the feed water quality and fouling potential. The high and low fouling experiments discussed here represent the highest and lowest flux decline recorded (at the end of the filtration) during the experimental period. The medium fouling event represents an experiment where the flux decline was between the highest and lowest fouling events. When the membranes were back-washed every hour, part of the flux decline was recovered. This recovery was considered to be indicative of the removal of the reversible membrane foulant components. The remaining residual flux decline was considered to represent the irreversible membrane foulant components that were tightly

attached to the membrane. The remaining part of the discussion will focus on the UF experiments presented in Figure 9-4.

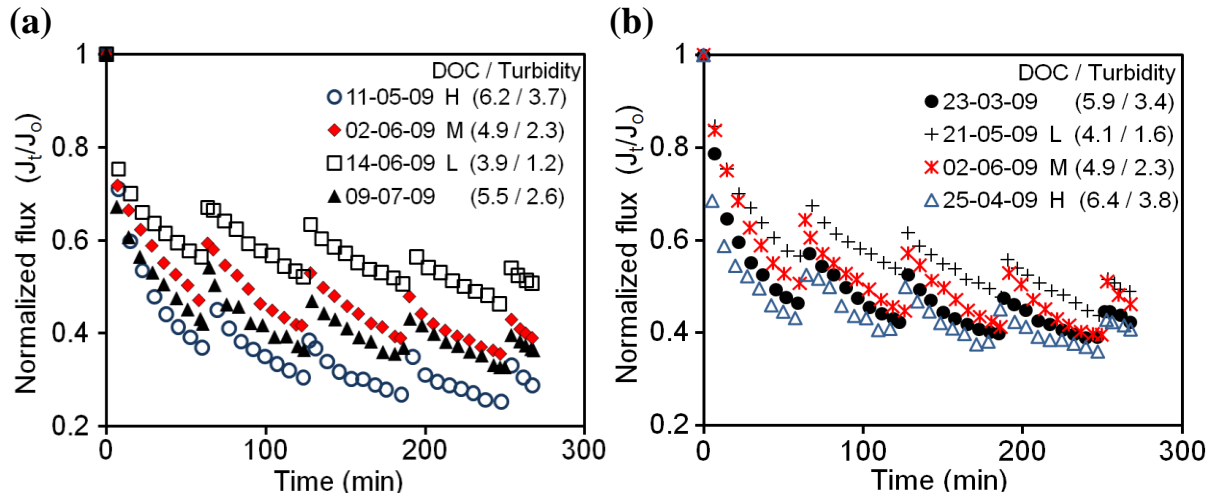


Figure 9-4 Normalized UF permeate flux profiles for (a) 60 kDa and (b) 20 kDa membranes with different feed water quality and fouling potential. DOC (mg/L) and turbidity (NTU) values of the membrane feed water are listed within parentheses. High, medium and low fouling experiments are identified as H, M and L, respectively.

The evolution of each normalized PC difference between the retentate and the permeate, $\Delta s_{1, \text{Norm}}$, $\Delta s_{2, \text{Norm}}$, $\Delta s_{3, \text{Norm}}$ and $\Delta s_{4, \text{Norm}}$, with respect to filtration time was examined to understand how the major foulant components contributed to the different fouling behaviour experienced during the filtration. Figure 9-5a, b, and c illustrate the accumulation of HS-like, colloidal/particulate and protein-like components, respectively for the 60 kDa UF experiments. Similarly, Figure 9-5e, f and g represent the accumulation of HS-like, colloidal/particulate and protein-like components, respectively for the 20 kDa UF experiments. In addition to Figure 9-5c, Figure 9-5d also illustrates the accumulation of colloidal/particulate matter for 60 kDa UF experiments. It should be noted that for every one hour filtration cycle, there are four $\Delta s_{j, \text{Norm}}$ data points for a given PC (or foulant component).

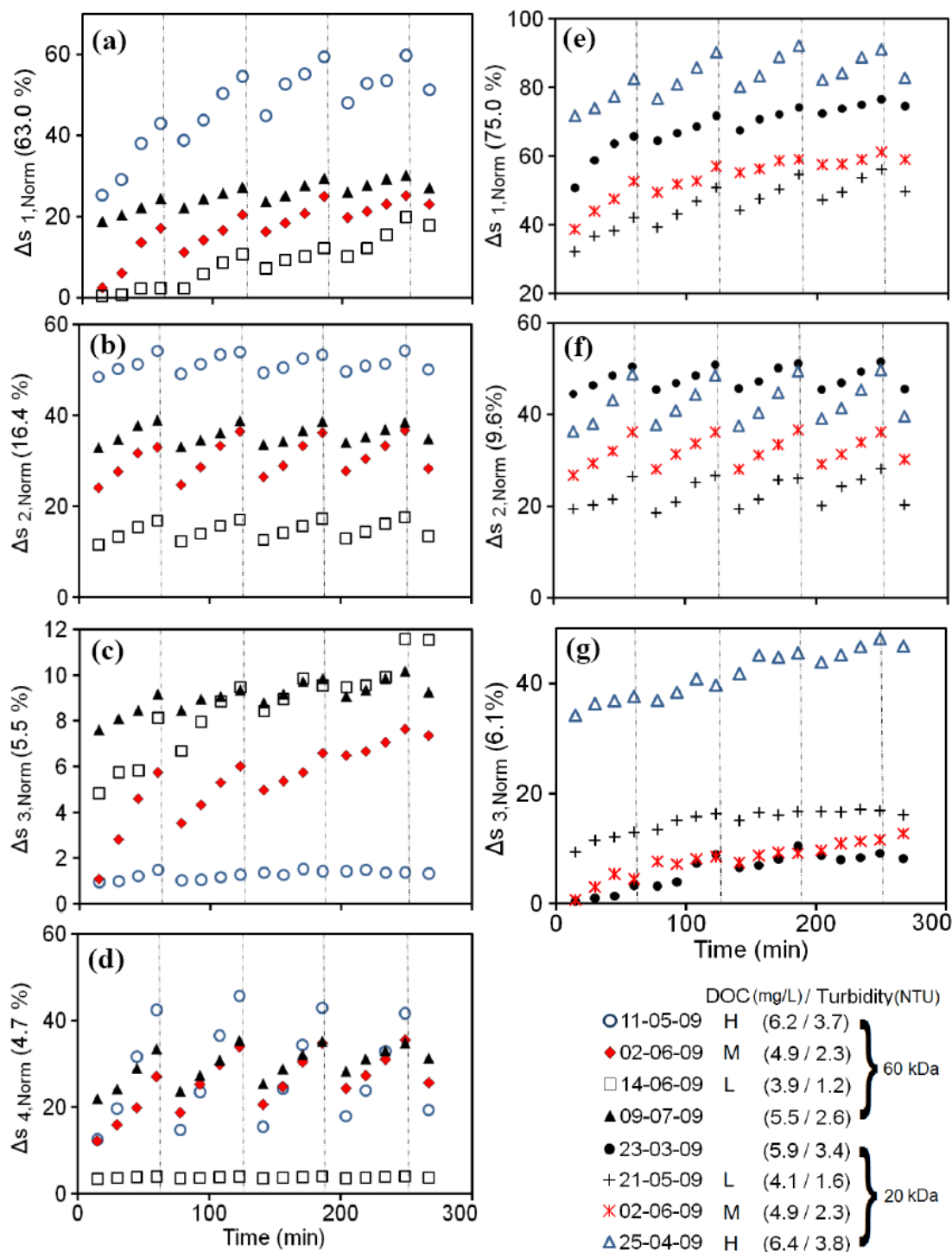


Figure 9-5 Plots of normalized PC score difference of retentate and permeate ($\Delta s_{j, \text{Norm}}$) vs. filtration time for 60 kDa (left: a, b, c and d) and 20 kDa (right: e, f, and g) membranes. Time of back-washing of the membranes is indicated by dashed vertical lines. $\Delta s_{1, \text{Norm}}$ and $\Delta s_{3, \text{Norm}}$ indicate the accumulation of HS- and protein-like material in and/or on the membrane, respectively. $\Delta s_{2, \text{Norm}}$ and $\Delta s_{4, \text{Norm}}$ indicate the accumulation of colloidal/particulate matter. High, medium and low fouling experiments are identified as H, M and L, respectively.

When examining the evolution of normalized permeate flux for the 60 kDa and 20 kDa UF experiments with respect to time (Figure 9-4) and the corresponding plots of normalized PC score difference between the retentate and the permeate, i.e. $\Delta S_{1, \text{Norm}}$, $\Delta S_{2, \text{Norm}}$, $\Delta S_{3, \text{Norm}}$ and $\Delta S_{4, \text{Norm}}$ (Figure 9-5), the following observations can be made:

- 1) An increase in the accumulation of colloidal/particulate components in and/or on the membrane was observed during the filtration cycles as illustrated in Figure 9-5b, d and f; however, the build-up of colloidal/particulate components appear to be more or less completely removed after each back-wash (i.e. for a given filtration, $\Delta S_{2, \text{Norm}}$ and $\Delta S_{4, \text{Norm}}$ values returned approximately to their initial levels in each permeation cycle after back-washing). This indicates that colloidal/particulate components are largely responsible for reversible membrane fouling for the 60 kDa and 20 kDa membranes employed in this study and with GRW (filtered) as feed. Similar observations have also been made in previous UF fouling studies (Jermann *et al.*, 2008b).
- 2) An increase in the accumulation in and/or on the membrane of the HS- and protein-like foulant components was observed during the 60 kDa UF experiments (Figure 9-5a, c). The back-wash step only partially removed these accumulated components (i.e. $\Delta S_{1, \text{Norm}}$, $\Delta S_{3, \text{Norm}}$ were reduced after back-washing but never returned to their initial conditions for a given filtration). Unlike colloidal/particulate materials, these components are not removed as effectively by back-washing. This indicates that for the 60 kDa membrane, HS- and protein-like components contributed more significantly to irreversible fouling compared to the colloidal/particulate components. Similar observations apply to the 20 kDa membrane as indicated by the $\Delta S_{1, \text{Norm}}$, and

$\Delta S_{3, \text{Norm}}$ values (Figure 9-5e and g). These observations are consistent with the UF fouling interpretations provided for river water foulant extracts (Jucker and Clark, 1994; Aoustin *et al.*, 2001); and model foulants (Jones and O'Melia, 2001; Jermann *et al.*, 2008b).

- 3) The removal by back-washing of the protein-like components that accumulated in and/or on the membrane was less efficient for the 20 kDa membrane than for the 60 kDa membrane (Figure 9-5c and g).
- 4) The extent of the permeate flux decline during the 60 kDa UF experiments shown in Figure 9-4 appears to be correlated with the level of HS-like and colloidal/particulate matter that accumulated in and/or on the membranes as presented in Figure 9-5a and b. Similar observations can be made for all the 20 kDa UF experiments except the UF experiment performed with GRW obtained on 25-04-09. When the extent of flux decline was plotted as a function of $\Delta S_{1, \text{Norm}}$ (accumulation of HS-like matter) and $\Delta S_{2, \text{Norm}}$ (accumulation of colloidal/particulate matter) separately, using both 60 kDa and 20 kDa UF experimental results, the highest correlation was observed for the accumulation of HS-like matter (results not shown).
- 5) The 20 kDa UF experiment performed with GRW (23-03-09), demonstrated a slightly lower permeate flux compared to the experiments performed with GRW (25-04-09). This slightly lower permeate flux is however not correlated with the magnitude of colloidal/particulate components accumulated in and/or on the membrane as indicated by $\Delta S_{2, \text{Norm}}$ values in Figure 9-5f (i.e. magnitude of $\Delta S_{2, \text{Norm}}$ for UF with GRW (23-03-09) is higher than UF with GRW (25-04-09)). Also, the accumulation of protein-like components during the UF experiment performed with GRW (25-04-09) is

significantly greater than for the other three filtrations (Figure 9-5g). Therefore, it is reasonable to assume that this higher accumulation of protein-like material may have contributed to the high fouling behaviour experienced during the UF experiment performed with GRW (25-04-09) compared to GRW (23-03-09).

- 6) Higher levels of protein-like components can be expected to increase the irreversible membrane fouling of UF membranes as explained previously. Increased levels of protein-like components can also interact with other fractions such as the polysaccharide-like matter contained in the colloidal/particulate components and increase the synergistic effect on UF membrane fouling (Lee *et al.*, 2006; Amy, 2008; Jermann *et al.*, 2008b; Susanto *et al.*, 2008). Due to these reasons, the presence of high protein-like matter in the case of the UF experiment performed with GRW (25-04-09) may have contributed to the observed higher fouling behaviour.
- 7) Unlike the HS-like and colloidal/particulate fouling components, the extent of accumulated protein-like components does not appear to be directly correlated with the degree of fouling observed for the 60 kDa and 20 kDa UF experiments (Figure 9-4, Figure 9-5c and g). This is expected due to very low concentration levels of protein-like matter present in GRW (Peiris *et al.*, 2010), although as stated above, protein has an indirect effect on fouling through interaction with colloidal components.

9.3.5 Fluorescence EEMs of Accumulated Membrane Fouling Components

Fluorescence EEMs of the foulants that accumulated in and/or on the membrane during the 60 kDa and 20 kDa filtrations are illustrated in Figure 9-6a - d. Specifically, the accumulated

foulants were extracted from the 60 kDa (Figure 9-6a and b) and 20 kDa (Figure 9-6c and d) UF experiments performed with GRW (02-06-09). Figure 9-6a and c illustrate fluorescence EEMs of loosely attached foulants for 60 kDa and 20 kDa UF membranes, respectively. Figure 9-6b and d show the fluorescence EEMs of ethanol extracted foulants for 60 kDa and 20 kDa UF membranes, respectively. The fouling behaviour observed during these UF experiments was previously illustrated in Figure 9-4 and Figure 9-5. The loosely attached foulants represent the foulants that contribute to reversible fouling. The ethanol extracted foulants represent the foulants that could not be removed by physical cleaning methods, i.e. material that was firmly attached to membranes and that could largely contribute to irreversible fouling behaviour of UF membranes.

It should also be noted that fluorescence EEMs of the ethanol extracted foulants demonstrated fluorescence features that are typical to the major NOM components present in natural water (discussed below). This indicates that there were no significant conformational changes of the NOM macromolecules that had altered the fluorescence EEMs. Also, the near neutral pH ($\sim 7.1 - 8.3$) of the ethanol extracted foulants would also have contributed to minimizing these conformational changes that would otherwise be significant at high or low pH levels due to changes in the charge density of NOM macromolecules (Henderson et al., 2009).

9.3.5.1 Colloidal/particulate Matter Fouling

Fluorescence EEMs of the loosely-attached foulants of the 60 kDa (Figure 9-6a) and 20 kDa (Figure 9-6c) membranes show high intensities in the RS regions (> 800 a.u.). In contrast, the

fluorescence EEMs of the ethanol extracted foulants for the 60 kDa (Figure 9-6b) and 20 kDa (Figure 9-6d) membranes did not exhibit any significant fluorescence intensity in the RS regions (related to the presence of colloidal/particulate substances as discussed earlier). Using a PC score analysis (Figure 9-5), it was demonstrated that colloidal/particulate substances are mainly contributing to reversible fouling for both membranes examined in this study. Therefore it could be concluded that colloidal/particulate matter largely contributes to reversible fouling and is mainly present in the form of loosely-attached foulants. Also, the presence of weak fluorescence regions (peak α and β) related to humic components and the presence of a smaller fluorescence peak (δ) related to protein-like components (Figure 9-6 a and c) indicate that colloidal/particulate matter constitute the main components of the loosely-attached foulants.

9.3.5.2 Fouling by HS-like Components

Figure 9-6a and c indicate that HS-like components were not significantly present in the loosely-attached foulants. However, fluorescence EEM peaks (α) and (β) in the form of “shoulders” in Figure 9-6b and fluorescence EEM peak (α) also in the form of a “shoulder” in Figure 9-6d show the presence of HS-like components in the ethanol extracted foulants for both membranes. These results imply that the contribution of HS for UF fouling was due to the strong attachment to the surface and/or inside the pores of the membrane and hence confirms the above PC score analysis in which HS was identified as a significant contributor for irreversible fouling.

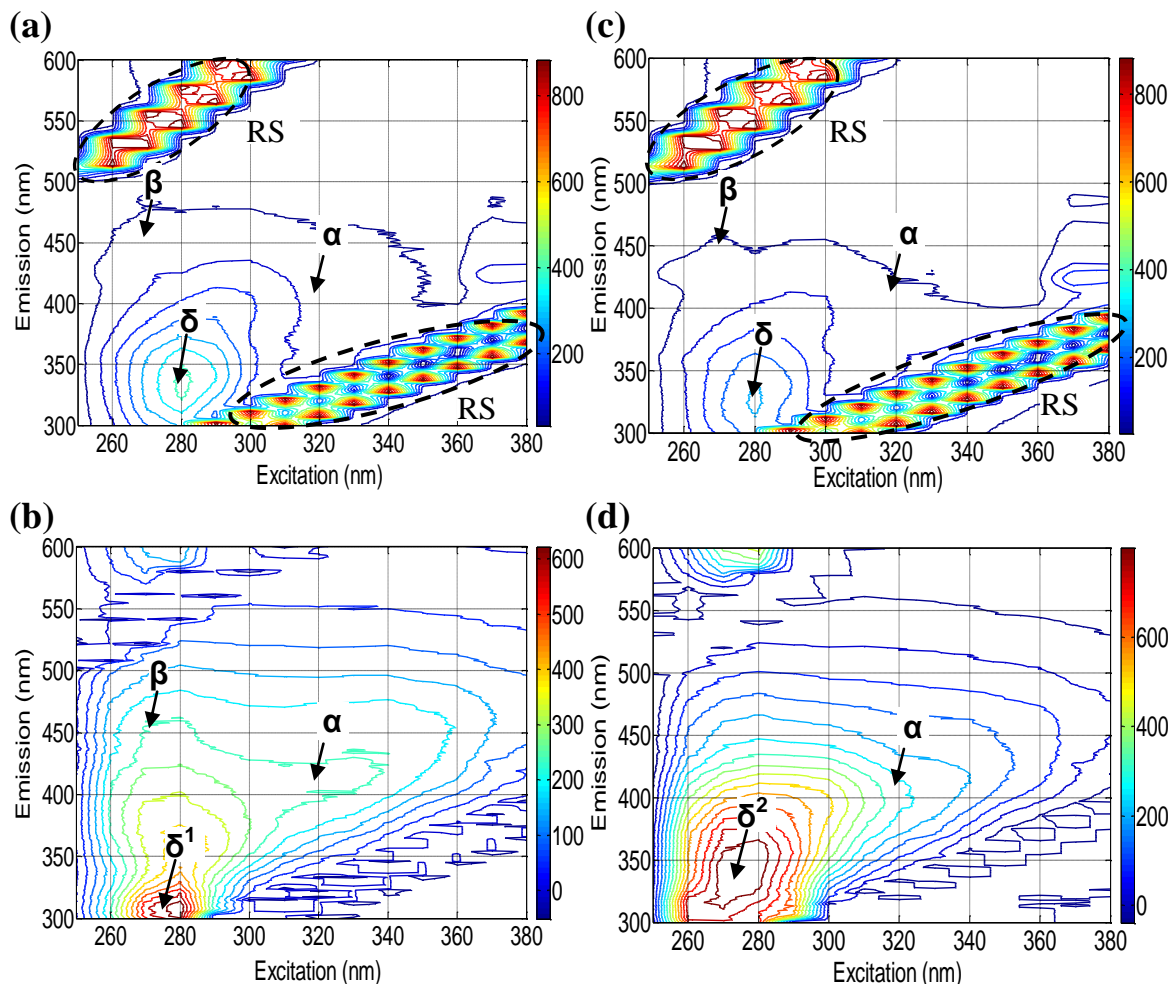


Figure 9-6 Fluorescence EEMs of the loosely-attached foulants (above) and ethanol extracted foulants (below) for 60 kDa (a and b) and 20 kDa (c and d) membranes. Rayleigh scattering (RS) regions are enclosed using dashed-lines. UF experiments were performed with GRW (obtained on 02-06-09).

9.3.5.3 Fouling by Protein-like Components

The presence of protein-like components in the ethanol extracted foulant corresponds to fluorescence EEM peaks (δ^1) and (δ^2) in Figure 9-6b and d. Even though these fluorescence EEM peaks (δ^1) and (δ^2) are located at slightly different excitation-emission wavelength coordinates, they remain located in a region that is typical for protein-like components (Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003). This indicates that protein-like components, as in the case of HS-like components, contributed to irreversible fouling of 60

kDa and 20 kDa membranes further corroborating the PC score analysis and fouling interpretation discussed above. In addition, peak (δ) in Figure 9-6a and c indicate the existence of protein-like components in the loosely attached foulants of both 60 kDa and 20 kDa UF membranes. Even though protein-like components were not identified as a main contributor for reversible fouling in the PC score analysis, protein-like components can be expected to be present in the loosely attached foulants due to their interactions with colloidal/particulate substances (Lee *et al.*, 2006; Amy, 2008; Jermann *et al.*, 2008b; Susanto *et al.*, 2008).

9.3.5.4 Different Membrane Rejection Characteristics

The differences in the fluorescence EEMs of the ethanol extracted foulants for the 60 kDa and 20 kDa UF membranes, as manifested in terms of (1) the disparity in peak positions of (δ^1) and (δ^2) and (2) the absence of a clear peak/shoulder (β) in the ethanol extract of the 20 kDa membrane (Figure 9-6b and d), are likely due to the differences in membrane rejection characteristics. The normalized flux decline recorded for both membranes with GRW (02-06-09) also illustrated this difference in the rejection characteristics (Figure 9-4a and b). This difference in the rejection characteristics of the 60 kDa and 20 kDa membranes were also apparent, when comparing the fluorescence EEM intensities at peak positions (α), (β), (δ^1) and (δ^2) for the ethanol extracted foulants of both membranes (Table 9-3). The ethanol extracted HS-like and protein-like components of the two membranes appear to be differently rejected as identified by the intensities of fluorescence EEM peaks (β), and (δ) (Table 9-3). In addition, a comparatively higher fluorescence EEM peak intensity related to protein-like components was identified in the ethanol extracted foulants of the 20 kDa membrane

compared to the 60 kDa membrane (Table 9-3). This corroborates the lower removal efficiency of the protein-like components during the back-washing of the 20 kDa membranes in comparison to the 60 kDa membranes as discussed earlier.

The UF experiments were performed at the same TMP for both membranes. Thus, the initial permeate flux was larger for 60 kDa membranes compared to 20 kDa membranes resulting in higher net convective transport of foulants for the 60 kDa membranes. This difference does not seem to have significantly contributed to higher irreversible fouling of 60 kDa membranes. In particular, irreversible accumulation of protein-like matter was higher for 20 kDa membranes compared to 60 kDa membranes as indicated by the higher fluorescence EEM peak intensity for protein-like components (Table 3). Nevertheless, this increased net convective transport of foulants with the 60 kDa membranes may explain the higher reversible fouling observed experimentally.

Table 9-3 Fluorescence EEM peak intensities of the ethanol extracted foulants for 60 kDa and 20 kDa membranes.

Foulant	EEM Peak intensity (a.u.)	
	60 kDa	20 kDa
HS -like: peak (α) (Ex:Em: 320/415)	243.9 \pm (4.8)	252.7 \pm (5.4)
HS-like (β) (Ex:Em: 270/460)	233.0 \pm (4.1)	207.0 \pm (4.5)
Protein-like (δ^1 , δ^2)	660.6 \pm (11.2) (Ex:Em: 280/310)	846.3 \pm (15.9) (Ex:Em: 270/340)

GRW used in the UF experiments were obtained on 02-06-09. Measurement errors with 95% confidence level (n= 3) are indicated within parentheses.

9.3.6 Potential for Process Monitoring and Future Applications

In this study, PCA of fluorescence EEMs proved suitable for characterizing three major membrane foulant fractions and capturing different fouling dynamics resulting from their accumulation on the surface and/or in the pores of UF membranes during filtration and back-washing cycles of GRW. The proposed PCA method has applications for assessing the contributions to membrane fouling by major membrane fouling components present in natural water and for understanding the fouling mechanisms of these components from a fundamental perspective. The proposed PCA-based method has the potential for monitoring membrane fouling and the identification of the predominant foulants present in natural water that contribute to reversible and irreversible fouling in a qualitative manner. Such monitoring capability may serve as an invaluable tool for the development of fouling control strategies to mitigate membrane fouling for the production of drinking water. Current research is now directed at using this approach to develop fluorescence-based mass-balance equations capable of quantitatively monitoring fouling behaviour associated with the UF of natural water. This approach may also be applicable for understanding fouling in other membrane treatment operations such as NF employed in drinking water applications.

9.4 Conclusions

A novel approach that combines PCA analysis and fluorescence EEM measurements to characterize three major membrane foulant fractions, identify different fouling behaviour observed during UF of natural river water and understand how the three major membrane

foulant fractions present in natural water contribute to reversible and irreversible fouling was presented. The following conclusions can be drawn:

1. The key membrane foulant components responsible for UF membrane fouling were identified as colloidal/particulate substances, HS- and protein-like matter. Each of these components was related to specific PCs generated by the PCA of fluorescence EEMs captured during the course of UF.
2. The accumulation of these major foulant components on the surface and/or in the pores of the membrane could be monitored by examining the temporal evolution of the PC score difference between the retentate and the permeate during UF of representative river water (GRW).
3. Colloidal/particulate components contributed mainly to reversible fouling of both 60 kDa and 20 kDa UF membranes. HS- and protein-like components contributed more significantly to irreversible fouling compared to colloidal/particulate materials. These observations were also confirmed by examining the fluorescence EEMs of the foulants extracted from the membranes.
4. The extent of permeate flux decline was generally correlated with the level of HS-like and colloidal/particulate components accumulated in and/or on the membrane while there was no direct dependence on fouling with respect to protein-like components, although increased levels appeared to contribute to high fouling behaviour due to their suspected interactions with colloidal/particulate components.

5. Fluorescence EEMs of the foulants extracted from the membranes revealed the different rejection characteristics of the 60 kDa and 20 kDa UF membranes investigated in this study.

CHAPTER 10

Fluorescence-based Fouling Prediction and Optimization of a Membrane Filtration Process for Drinking Water Treatment

Overview

A novel fluorescence-based approach is proposed for modeling and predicting different fouling dynamics in an ultrafiltration (UF) process for drinking water treatment. Principal component analysis (PCA) was used to extract principal components (PCs), related to major membrane foulant groups, from fluorescence excitation-emission matrix measurements captured during the UF of natural river water using a bench-scale membrane cross-flow set-up. The evolution of the PC scores over the filtration time was then related to membrane fouling by using PC score balanced-based differential equations. This approach was suitable for forecasting fouling behaviours with good accuracy based solely on fluorescence data collected 15 minutes into the filtration process. The proposed approach was tested experimentally as a basis for optimization by modifying the UF back-washing times with the objective of minimizing energy consumption and maximizing water production. This approach was also useful for identifying the fouling groups contributing to reversible and irreversible membrane fouling.

Keywords: drinking water treatment, fluorescence spectroscopy, modeling membrane fouling, principal component analysis, real-time optimization

10.1 Introduction

Membranes are widely used in drinking water applications to achieve different treatment objectives such as improved removal of colloidal/particulate matter, pathogenic organisms, natural organic matter (NOM) and salinity in water. Different types of membrane systems such as microfiltration, ultrafiltration (UF), nanofiltration and reverse osmosis are being increasingly used individually or in combination (hybrid mode) to accomplish these treatment objectives and to produce drinking water with consistent quality (Fiksdal and Leiknes, 2006). Membrane-based technology also allows a smaller footprint for the treatment facilities compared to conventional treatment processes (Atkinson, 2004). However, membrane fouling, which is the result of the accumulation of materials (foulants) on the surface and/or in the pores of the membranes, is a major constraint when considering both the adoption and performance consistency of membrane-based treatment operations. NOM fractions, such as humic substances (HS), protein- and polysaccharide-like substances as well as colloidal/particulate matter present in water, are mainly responsible for membrane fouling in drinking water applications (Saravia *et al.*, 2006; Jermann *et al.*, 2007). In practise, membrane fouling is controlled by implementing cleaning operation schemes that include membrane back-washing (also known as back-flushing) (Ebrahim, 1994) and chemical cleaning of fouled membranes (Kimura *et al.*, 2005).

Fouling increases operational costs as a result of permeate flux decline and/or increased energy consumption due to higher trans-membrane pressure (TMP) requirements needed as the driving force for the production of drinking water. In addition, frequent chemical cleaning of fouled membranes leads to rapid deterioration of membrane performance,

shortened service life and increased costs. The efficient use of fouling controlling strategies can reduce the energy demand and other operational costs associated with fouling and improve the sustainability of membrane-based drinking water treatment operations to ensure high production of water. This can be accomplished by optimizing the operation of membrane filtration processes (Seidel and Elimelech, 2002).

10.1.1 Methods Available for Membrane Fouling Modeling

Optimization of a membrane filtration operation requires a model that is capable of predicting the extent of reversible and irreversible membrane fouling for given raw water feed. Many studies have focused on achieving this objective by assessing and/or predicting membrane fouling behaviour using mechanistic modeling approaches (Bowen *et al.*, 1995; Tansel *et al.*, 2000; Chang and Benjamin, 2003; Taniguchi *et al.*, 2003; Bolton *et al.*, 2006). These studies however focused mainly on dead-end and a few cross-flow membrane filtration systems that did not involve membrane back-washing cycles which are typically applied in drinking water treatment systems. Also, these modeling approaches are not suitable for successfully predicting membrane fouling in drinking water applications.

Modeling methods referred to as empirical or black-box approaches such as artificial neural networks (Delgrange-Vincent *et al.*, 2000; Cabassud *et al.*, 2002), empirical models (Shengji *et al.*, 2008) and genetic programming (Lee *et al.*, 2009) have also been used to correlate membrane fouling with long-term membrane feed water quality parameters and operation data such as turbidity, temperature, dissolved organic carbon content (DOC) and TMP in pilot scale filtration studies. These models are not able to capture the changes in the different

membrane foulant fractions in water during the filtration nor can they relate different fouling behaviour to individual membrane foulant fractions. As a result, since the individual relationships between the input variables and the predicted membrane flux are not developed based on engineering criteria, the successful implementation of optimization strategies for fouling control based on these black-box models is not always warranted. In addition, from a membrane research stand point that is geared towards improving membrane fouling characteristics, the above mentioned black-box techniques are not suitable for relating the degree of fouling to the relative concentrations of NOM and other foulant components present in water and are not helpful in addressing remedies for controlling fouling.

10.1.2 Proposed Fouling Modeling and Optimization Approach

This study proposes a fluorescence-based modeling approach that is capable of capturing the dynamic changes of different membrane foulant fractions that occur during the UF of natural water for the production of drinking water. This model is developed based on fluorescence excitation-emission matrix (EEM) measurements made during UF operation to characterize different membrane foulant components present in water.

The fluorescence EEM method captures specific fluorescence features that correspond to HS- and protein-like materials (Baker, 2001; Henderson *et al.*, 2009) as well as particulate/colloidal matter (Peiris *et al.*, 2010) present in water. The fluorescence EEMs capture a large number of intensity readings recorded at different excitation and emission wavelengths for natural water samples. Compared to other available NOM membrane foulant characterization methods (Huber *et al.*, 1992; Her *et al.*, 2003; Gray *et al.*, 2007), this

approach is capable of differentiating the major membrane foulant fractions and is suitable for performing rapid, direct and accurate analysis with high instrumental sensitivity (Coble *et al.*, 1990; Peiris *et al.*, 2008).

In a previous study, principal component analysis (PCA) was successfully used to deconvolute spectral information captured within fluorescence EEMs into principal components (PCs) that were related to HS, protein-like and colloidal/particulate matter present in natural water (Peiris *et al.*, 2010). This PC score-based approach is suitable for rapid monitoring of the performance of a membrane-based drinking water treatment system with high sensitivity. The same approach was therefore used in this study to generate PC scores that correspond to the fluorescence EEMs captured over the course of the UF filtration operation that contained cycles of permeation and membrane back-washing. The novelty of the current work as compared to the previous study is that these PC scores were then used as states within a system of differential equations representing approximate mass balances of the main foulant groups, i.e. HS, protein-like and colloidal/particulate matter. Thus, the resulting mathematical model can be viewed as a hybrid one where dynamic balances are performed over scores obtained from a PCA of experimentally obtained fluorescence data. Based on the UF fouling dynamics predicted by this modeling approach, an optimization strategy is proposed for the estimation of the optimal membrane back-washing scenario that minimizes energy consumption while maximizing drinking water production. The proposed modeling approach is able to forecast/predict membrane fouling behaviour and thus it is useful for the optimization of membrane filtration operations. The potential of this approach

as a membrane research tool by providing an understanding of how major membrane foulant fractions contribute to reversible and irreversible fouling behaviour is also discussed.

10.2 Materials and Methods

Grand River water (GRW) (Southwestern Ontario, Canada) was filtered using a 200 micron filter (038A-2080; Keller Products, Inc. Acton, MA) and used as the feed in UF experiments. The DOC of the feed ranged from 3.9 – 6.5 mg/L and its turbidity values were in the range of 1.2 – 3.8 NTU during the experimental period (March – November 2009). GRW was stored at 4 °C before the experiments and always used within 48 hours of the collection time.

10.2.1 Bench-scale Membrane Filtration Set-up

UF experiments were conducted at constant TMP using a bench-scale flat sheet cross-flow set-up as shown in Figure 10-1. The membrane cross-flow cell holder (Sterlitech CF042; Sterlitech Corporation, Kent, WA) used in this study had an effective membrane area of 42 cm². Flat sheet UF membranes with a molecular weight cut-off (MWCO) size of 20 kDa (Polysulfone - YMEWSP3001; GE Osmonics) and 60 kDa (Polyethersulfone - YMPWSP3001; GE Osmonics) also from Sterlitech Corporation were used. Contact angle measurements performed on virgin 20 kDa and 60 kDa membranes were $72 \pm 2^\circ$ (n = 6) and $80 \pm 1^\circ$ (n = 6), while the initial pure water flux at TMP = 15 psi (103.4 kPa) were ~1.6 and ~2.4 L/min.m², respectively. A new membrane was used for each filtration run and prior to the start of each experiment, the membranes were compacted at 15 psi using Milli-Q water until a stable permeate flow was achieved.

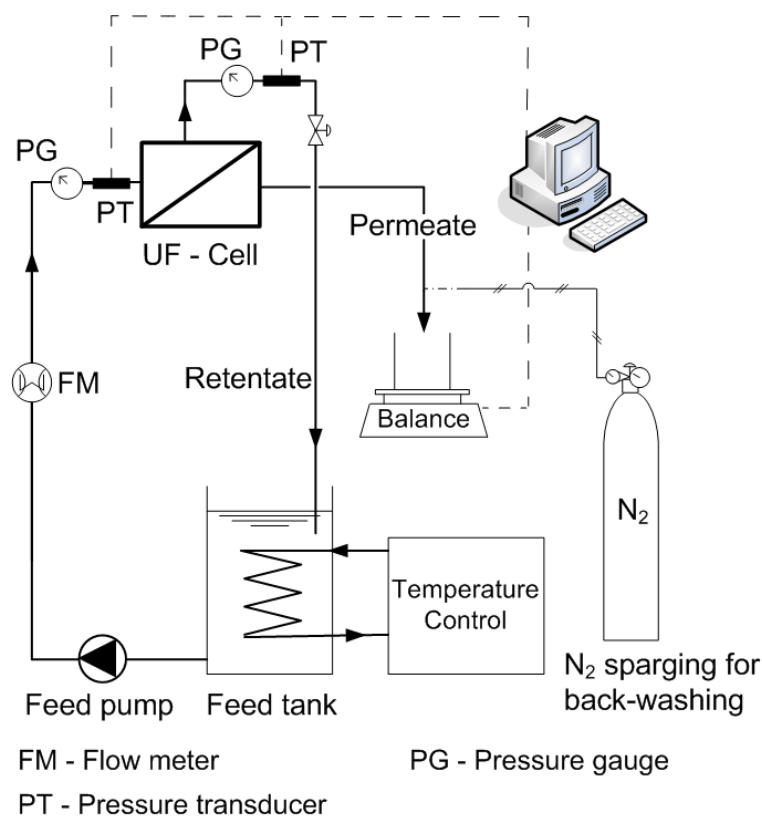


Figure 10-1 Bench-scale ultrafiltration cross flow set-up.

Feed water to the membrane set-up was maintained at 0.6 L/min with a TMP of 15 psi. Retentate was circulated back to the feed tank which contained 22 L of water that was maintained at $\sim 25 \pm 1^\circ \text{C}$ using a temperature controller. The permeate water was continuously removed and its mass and corresponding permeate flux was recorded using a balance connected to a computer using a LabView-based (LabView 8.0; National Instruments, Austin, TX) interface. The filtration consisted of a two step operation cycle: (1) permeation period and (2) back-washing for 20 s. For non-optimized conditions, the permeation period was 1 h while for optimized back-washing the permeation period was adjusted according to the back-washing times determined from the solution of an optimization problem as will be discussed later. Back-washing of the membrane was implemented by forcing the permeate (which is the liquid in the permeate pipe and permeate

channels of the membrane cell holder) in the opposite direction through the membrane using Nitrogen gas at 10 psi (68.9 kPa). During the back-washing, feed flow was maintained over the membrane surface to induce shear force on the membrane surface thus assisting in the removal of foulants. Fluorescence EEMs of both retentate and permeate were recorded at 15 min intervals during the course of the filtration.

10.2.2 Fluorescence Analysis

The fluorescence analysis procedure explained in Peiris *et al.* (2010) was used in this study to record fluorescence EEMs using a Varian Cary Eclipse Fluorescence Spectrofluorometer (Palo Alto, CA). A detailed description of the fluorescence EEM analysis procedure, methods used in fluorescence signal correction and the selection of the spectrofluorometer parameter settings used in this study for obtaining reproducible fluorescence signals, is found in Peiris *et al.* (2008, 2009 and 2010). During the course of the fluorescence analyses, there were no significant differences in Raman scattering peak intensities recorded for Milli-Q water at Ex/Em ~ 348 nm/396 nm (i.e. difference was less than 2%), confirming that there were no significant fluctuations in the performance of the spectrofluorometer lamp or other hardware. The temperature of the water samples were maintained at room temperature (~ 25 °C) during the analysis. The pH of all the water samples did not change significantly (pH ~ 7.8 – 8.4) during the experiments and no pH adjustment was made prior to the fluorescence analysis as the fluorescence EEMs are not significantly affected due to small pH differences (Spencer *et al.*, 2007).

10.2.3 Other Analytical Methods

The methods used for obtaining DOC, turbidity and membrane contact angle measurements are as previously described (Peiris et al., 2008 and 2010).

10.2.4 Fluorescence Data Pre-treatment and PCA

The fluorescence EEM of each sample contained 4214 excitation and emission coordinate points. The fluorescence intensity values corresponding to all 4214 coordinate points (spectral variables) of each EEM were rearranged following the fluorescence EEM data rearrangement procedure described by Peiris *et al.* (2010). This resulted in a $n \times 4214$ fluorescence data matrix, with each row containing fluorescence EEM data points of each sample; n represents the total number of samples composed of both retentate and permeate samples obtained during the UF experiments as described above. This procedure was followed in order to generate two data matrices referred to as matrix X60 and matrix X20 for UF experiments performed with 60 kDa and 20 kDa membranes, respectively. X60 and X20 data matrices contained 525 and 560 fluorescence EEMs from 15 and 16 different UF experiments, respectively. PCA was performed on matrix X60 and X20 separately to generate PC scores as explained in Persson and Wedborg, (2001) and Peiris *et al.* (2010).

PCA is generally used to extract a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and mathematically represented by linear combinations of original variables in the X matrix (X60 or X20 matrix in this case). These new variables, referred to as PCs, are able to describe major trends in the original spectral data sets (Peiris *et al.*, 2010) of X60 and X20. PCA decomposes the data matrix X as the sum

of the outer product of vectors \mathbf{s}_i and \mathbf{p}_i plus a residual matrix E as presented in Equation 10.1.

$$X = \sum_{i=1}^n s_i \cdot p_i + E \quad (10.1)$$

The \mathbf{s}_i vectors are known as scores (i.e. values) on the PCs (i.e. new variables) extracted by PCA. The \mathbf{p}_i vectors are known as loadings and contain information on how the variables (fluorescence variables in this case) relate to each other. By examining the loading values related to each PC, it is possible to understand which original spectral variables in the X matrix are better explained by each PC. A more detailed description about PCA can be found in Wold *et al.* (1987) and Eriksson *et al.* (2001). Before performing PCA analysis, both X60 and X20 matrices were auto-scaled, i.e. adjusted to zero mean and unit variance by dividing each column by its standard deviation. To determine the number of PCs that were statistically significant in capturing the underlying features in the X60 and X20 data sets, the leave-one-out cross-validation method (Eriksson *et al.*, 2001) was implemented. All computations were performed using the PLS Toolbox 5.2 (Eigenvector Research, Inc., Manson, WA) within the MATLAB 7.8.0 computational environment (MathWorks, Natick, MA).

10.3 Theoretical Basis

The statistically significant PCs calculated as explained above were found to be correlated to different membrane foulants such as HS-like, protein-like and particulate/colloidal matter

present in water (explained later). The evolution of these PC scores (s_i) corresponding to the principal components and consequently to the foulant fractions present in water, was therefore expected to be related to the membrane fouling behaviour as demonstrated by Peiris *et al.* (2010). The PC scores (s_i) associated with the retentate and permeate of UF processes were therefore used to formulate a model of the fouling behaviour experienced by 60 kDa and 20 kDa membranes.

10.3.1 PC-based Modeling of Membrane Fouling

Since it is impossible at this point to quantify each of the individual species in natural water that contribute to fouling it was decided instead to perform a balance on the PC scores which are representative of different groups of foulants, i.e. HS-like, protein-like and colloidal/particulate. Thus, the accumulation of membrane foulants on the surface and in the pores of the membrane was calculated based on the PC score balance for a given group of foulants. These mass balances, performed on the control volume of the solution occupied by the membrane, were formulated as a way to account for the mass balances of the individual foulant species present in the water. Accordingly, the accumulation of the membrane foulant (j) that contributes to fouling can therefore be represented as follows:

$$\frac{ds_{j,M}}{dt} = \frac{1}{kV_M} \left[(1-w)A \frac{\Delta P}{\mu R_t} (s_{j,R} - s_{j,P}) - wL_j \right] \quad (10.2a)$$

$$L_j = \dot{m}_{wash} eff_j e^{-q_j R_t} s_{j,M} \quad (10.2b)$$

for $j=1, 2, 3, \dots, N$ and $w = 0$ or 1

Where s_j is the PC score related to the j^{th} membrane foulant at time = t. N is the number of PCs generated by PCA which are statistically significant and deemed to be important for capturing the information related to the major groups of foulants. Subscripts R, P and M denote retentate, permeate and the membrane, respectively. V_M is the volume of the solution occupied by the membrane and k is a parameter that specifies the active portion of V_M (i.e. actual portion of V_M that participates in the filtration). The effective membrane surface area, TMP and the water viscosity are represented by symbols A, ΔP and μ respectively. \dot{m}_{wash} is the volume flow rate used for periodic membrane back-washing, w is a binary variable that models permeation through the membrane ($w = 0$) or back-washing ($w = 1$). eff_j represents the efficiency at which the j^{th} foulant fraction (i.e. foulant fraction related to j^{th} PC) was removed during the back-washing. q_j is a parameter describing the decay of efficiency in back-washing over time due to irreversible fouling caused by the j^{th} membrane foulant. Irreversible fouling is attributed to the accumulated membrane foulant material that cannot be removed by membrane back-washing. R_t is the membrane resistance at time = t, which is given in terms of the scores as follows:

$$R_t = R_0 + \sum_{j=1}^N \beta_j s_{j,M} + \beta_{inter} s_{protein,M} \times s_{coll./partic,M} \quad (10.3)$$

R_0 is the initial membrane resistance before fouling occurs. β_j , $j=1, 2, 3, \dots, N$ are the model parameters. β_{inter} is also a model parameter related to the interaction between protein and colloidal/particulate matter (represented by $s_{protein,M}$ and $s_{coll./partic,M}$ respectively) that contributes to membrane fouling. The existence of this interaction was found to be significant

in a separate correlation analysis study (results not shown) and found to be very important for improving the model predictions in this study. It should be noted that while the values of the scores (i.e. s_j values) change with time as per the differential Equation 10.2a, the β 's are regressed off-line and do not change with time.

Also, the processes involved in the transfer of membrane foulants from the retentate to the membrane or vice versa are quite complex involving deposition of foulants due to attractive forces and removal due to shear stresses acting on the foulant layers. Detailed modeling of these phenomena is difficult. Therefore it was assumed that the net amount of foulant transfer from the retentate to the membrane is equal to the accumulation of foulants on the surface and in the pores of the membrane as follows:

$$D_j(s_{j,R} - s_{j,M}) = \frac{\Delta P}{\mu R_t} (s_{j,R} - s_{j,P}) \quad (10.4)$$

Where D_j is the effective diffusivity coefficient of the j^{th} foulant fraction. D_j is a lumped parameter that combines all possible mass transfer mechanisms involving the transfer of membrane foulants from the retentate to the membrane or vice versa as mentioned above.

The permeate water flux through the membrane at time = t is as follows:

$$J_t = \frac{\Delta P}{\mu R_t} \quad (10.5)$$

10.3.2 Model Calibration and Validation

Experimental permeate water flux data obtained by UF runs performed using GRW with different DOC content and turbidity values within the ranges as indicated above were used to calibrate the state space model given by the system of Equations 10.2a, 10.2b, 10.3, and 10.5. The model calibration involved the estimation of the model parameters k , β_1 , β_2 , β_3 , ..., β_N , $\beta_{\text{inter.}}$, eff_1 , eff_2 , eff_3 , ..., eff_N and q . This was achieved by minimizing the sum of squares error (SSE) between experimental and model estimations of permeate water flux by using the MATLAB function “ga”, a genetic algorithm code available within the MATLAB 7.8.0 computational environment. Model parameters, estimated in separate model calibrations for 60 kDa and 20 kDa UF membranes, are listed in the Supplementary Information (SI) Table S - 10-1. The model estimations were generated using the PC scores of retentate ($s_{j,R}$) and permeate ($s_{j,P}$) that correspond to fluorescence EEM measurements, obtained every 15 min during the course of UF. These scores were used as inputs to Equations 10.2a and 10.2b whereas the output was the corresponding score value at the membrane $s_{j,M}$ calculated from Equation 10.2a. The MATLAB ordinary differential Equation (ODE) solver “ode23” was used in solving the above state space model. Model validation was achieved using additional experimental permeate water flux and fluorescence EEMs data that were not used in the calibration. UF experimental data with low, medium and high fouling events involving data from a total of 9 and 10 experiments for 60 and 20 kDa UF membranes, respectively, were used for model validation.

10.3.3 Generation of Model Predictions

The model given by the system of Equations 10.2a, 10.2b, 10.3, 10.4 and 10.5 was then used to obtain model predictions based solely on the fluorescence EEMs of retentate and permeate captured at time = 15 min after the start of the UF experiments. PC scores (S_{j,R_15min} and S_{j,p_15min}) that are related to these fluorescence measurements were used for the estimation of the predicted permeate water flux into the future along a total time horizon of 4 h. During the calculations of model predictions, the PC scores related to the retentate were assumed to be constant and equal to the values obtained at time = 15 min during the prediction period (i.e. $S_{j,R} = S_{j,R_15min}$ in Equation 10.4). This assumption was based on the very small changes (< 5% increase in most cases) in the PC scores of retentate observed during the UF experiments. Also, the initial estimations of the effective diffusivity coefficient D_j ($j= 1, 2, 3, \dots, N$) was calculated using Equation 10.4 based on PC scores corresponding to the fluorescence EEMs of retentate and permeate captured at time = 15 min. These initial estimations were subsequently updated according to the following equation during the calculation of model predictions to account for the change of D_j resulting from membrane fouling over time. This was deemed to be necessary for approximating the evolving fouling conditions over time.

$$D_{j,t} = z_1 D_{j,int} + z_2 D_{j,t-\Delta t} \quad (10.6)$$

for $j=1, 2, 3, \dots, N$

Where $D_{j,t}$ is the effective diffusivity coefficient of the j^{th} foulant fraction at time $t = t$, $D_{j,int} =$ is the initial estimate of D_j and $D_{j,t-\Delta t}$ is the value of D_j at time $t = t - \Delta t$. Δt is the constant time step length ($\Delta t = 1$ s) used by the ODE solver. Z_1 and Z_2 are parameters that were

estimated by minimizing the SSE between model predictions and measured permeate water flux using a genetic algorithm approach as mentioned above. It should also be noted that Equation 10.6 does not necessarily cause D_j to increase over time. As the accumulation of foulant content in the membrane increases, the removal of foulants from the membrane to the retentate becomes significant, causing D_j to decrease as per Equation 10.4. The prediction ability of the model was also validated with additional experimental permeate water flux and fluorescence EEM data that were not used in estimating the Z_1 and Z_2 parameters.

10.3.4 Optimization of the UF Process

The predicted permeate water flux can be used to understand the extent of fouling of the membrane and the reduced permeate water flux occurring over time for constant TMP operations (as demonstrated in this study). On the other hand, if constant permeate flux is desired, the TMP would increase as a result of fouling. In both situations, membrane fouling results in an increase in the energy requirement per unit amount of drinking water produced.

In this study, the UF membrane back-washing times were used as optimization variables to optimize the UF process so that the energy requirement per unit amount of drinking water produced was minimized. This optimization approach was implemented by minimizing the following objective function (OF), (Equation 10.7) subjected to the constraints listed in Equations 10.10 and 10.11.

$$OF = \frac{\text{Energy consumption}}{\text{Water production}} \quad (10.7)$$

Where energy consumption and the water production for time duration = Δt is given by:

$$\text{Energy consumption} = \frac{A(\Delta P)^2 \Delta t}{\mu R_t} \quad (10.8)$$

$$\text{Water production} = J_t A \Delta t \quad (10.9)$$

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} \leq \begin{bmatrix} -t_w \\ -t_w \\ -t_w \\ t_F - t_w \end{bmatrix} \quad (10.10)$$

$$t_1 \geq t_d \quad (10.11)$$

Where t_1 , t_2 , t_3 and t_4 are the times at which the back-washing of the UF membrane was implemented. The number of back-washing cycles was limited to four as this was sufficient to demonstrate the application of the proposed approach. The number of back-washes represent another parameter that could be included in this optimization approach but this will be addressed in future research. Also, $t_w = 180$ s is the sum of the time for back-washing (20 s) and the time required to connect and disconnect the Nitrogen gas supply for back-washing and for adjusting the TMP of the UF membrane cell holder (160 s), which were performed manually. The total filtration time is indicated by t_F (= 257 min) and t_d (= 15 min) is the time at which the first set of fluorescence EEMs of the retentate and permeate for UF operation were collected. This information was required for the model predictions as explained in

Section 10.3.3. The minimization of the OF (Equation 10.7) subjected to the constraints (Equations 10.10 and 10.11) were performed by using the MATLAB function “ga”.

10.4 Results and Discussion

10.4.1 Typical Fluorescence Spectral Features of Grand River Water

The fluorescence EEMs of GRW used in this study contained fluorescence regions that are representative of the presence of major membrane foulants such as HS- and protein-like NOM (Peiris *et al.*, 2010). These results are also consistent with fluorescence EEM data published elsewhere (Coble *et al.*, 1990; Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003; Sierra *et al.*, 2005). In addition, Rayleigh scattering regions (RS) observed in the fluorescence EEMs of GRW can also provide information related to the particulate/colloidal matter (Peiris *et al.*, 2010). Typical fluorescence EEM spectra of GRW – similar to the ones recorded in this study – that indicate these regions have been previously reported (Peiris *et al.*, 2008 and 2010) and are therefore not presented here but are included in SI Figure S - 10-1.

10.4.2 PCA of Fluorescence Data

PCA of X60 and X20 matrices generated four and three statistically significant PCs, respectively, capturing nearly 90% of the total variance present in the original spectral variables obtained from 60 kDa and 20 kDa UF experiments (Table 10-1). These PCs were found to be related to different membrane foulant fractions present in water as shown in Table 10-1. This was verified by examining the loading plots corresponding to each PC,

(generated from the loading values, i.e. $\mathbf{p_i}$ values) as demonstrated by Peiris *et al.* (2010). For example, the loading peak of PC – 1 appeared in the same location where the fluorescence EEM regions related to HS-like NOM. Similar observations were made with PC – 2, PC – 3 and PC – 4 in relation to the foulant fractions they represent as indicated in Table 10-1. The loading plots of each PC corresponding to 60 kDa and 20 kDa UF experiments can be found in SI Figure S - 10-2 and Figure S - 10-3. The remaining variance (~ 10%) in each case was considered to be due to the combination of unexplained variance by these PCs and the instrumental noise (determined to be less than 5% of the intensity readings) in the fluorescence measurements. Although it is possible to capture the remaining variance by generating additional PCs, additional PCs were not found to be related to any major membrane foulant fractions present in water (see Table 10-1).

Table 10-1 Variance captured in the PCA of the X60 and X20 matrices.

Principal component	X60 matrix		X20 matrix	
	(60 kDa UF spectral data)		(20 kDa UF spectral data)	
	Variance captured (%)	Related membrane foulant	Variance captured (%)	Related membrane foulant
1	63.0	Humic substances	75.0	Humic substances
2	16.4	Colloidal/particulate	9.6	Colloidal/particulate
3	5.5	Protein-like	6.1	Protein-like
4	4.7	Colloidal/particulate	-	-
Total	89.6		90.7	

10.4.3 Model Predictions

Model predictions and the experimentally measured normalized permeate water flux of selected 60 kDa and 20 kDa UF experiments that were not used in the model calibration/parameter estimation are shown in Figure 10-2. Permeate flux was normalized with respect to initial pure water flux of the membranes. These experiments correspond to low, medium and high membrane fouling situations. The sudden increase in permeate flux occurring immediately after the exponential-type flux declines of the permeation step corresponds to the membrane back-washing. The model predictions for these experiments were obtained using only the fluorescence-based PC scores of retentate and permeate obtained at time = 15 min of the UF as explained earlier. The permeate flux prediction results indicate that the model was able to successfully predict different membrane fouling behaviours experienced by both UF membrane types.

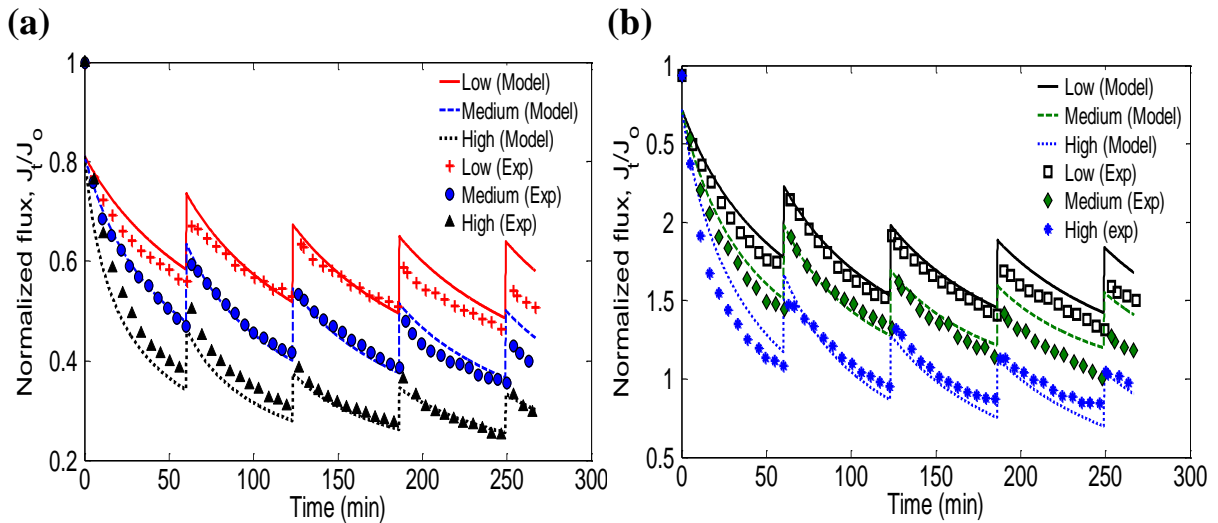


Figure 10-2 Model predictions (lines) and experimentally measured (symbols) normalized permeate water flux for selected (a) 60 kDa and (b) 20 kDa UF experiments of low, medium and high membrane fouling situations.

These results indicate that the proposed fouling modeling approach can be used for forecasting different fouling behaviours corresponding to changes in the membrane feed water quality. Hence, this modeling approach can be used to detect high membrane fouling events well in advance and thus appropriate process optimization measures could be implemented to ensure sustainable operation of drinking water treatment systems.

10.4.4 Optimization of UF for Drinking Water Treatment

The proposed modeling approach was then applied for the optimization of 60 kDa and 20 kDa UF operations as explained in Section 3.4. GRW obtained on Oct. 25 and Nov. 23, 2009, was used in the 60 kDa and 20 kDa experiments, respectively.

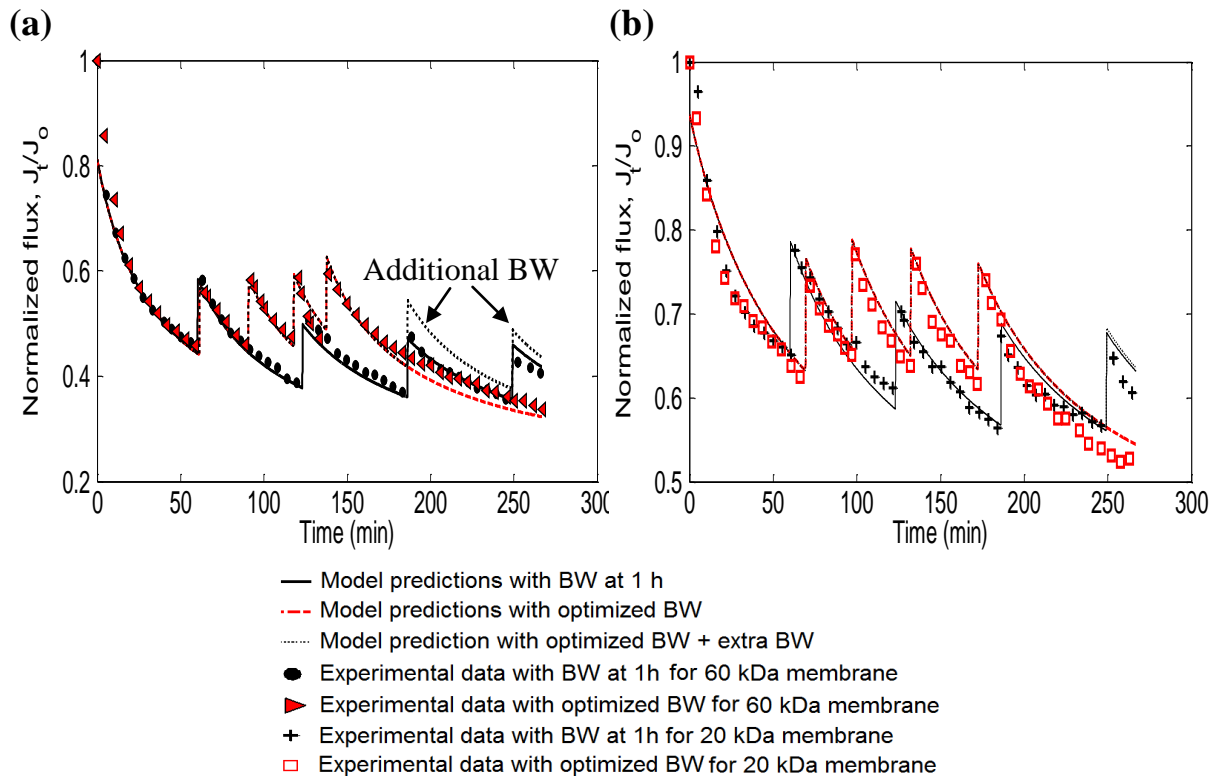


Figure 10-3 Model predictions (lines) and experimentally measured (symbols) normalized permeate water flux obtained for (a) 60 kDa and (b) 20 kDa UF operations with normal back-washing (BW) times (every hour) and optimized back-washing times.

Figure 10-3a and b demonstrate the model predictions of the fouling behaviour for UF of GRW (pre-filtered) using 60 kDa and 20 kDa membranes with back-washing at regular time intervals of 1 hr (i.e. before implementing optimal backwashing intervals). When back-washing times were optimized using the proposed optimization approach (Section 10.3.4) for the 60 kDa membrane, the model predictions indicated an energy savings of 3.7% with a 4.3% increase in the total volume of drinking water production. The back-washing times generated by the optimization approach were $t_1 = 61$ min, $t_2 = 90$ min, $t_3 = 118$ min, and $t_4 = 137$ min. Applying the same approach for the 20 kDa membrane, it was possible to achieve energy savings of 2.6% with a 3.1% increase in the total volume of water production. The corresponding optimum back-washing times for the 20 kDa membrane were $t_1 = 69$ min, $t_2 = 97$ min, $t_3 = 132$ min, and $t_4 = 172$ min. The optimal backwashing cycles were implemented experimentally in order to test the validity of the optimization results. The results of these experiments are shown in Figure 10-3a and b indicating good agreement between experimental values and model predictions thus confirming the optimization solution.

Although the optimization approach developed in this study was limited to four back-washing cycles, it is possible to further improve the energy savings and the water production by employing additional back-washing cycles with the optimized conditions. For example, when two additional back-washing cycles were included as illustrated in Figure 10-3a, the model predictions indicated an increase in the energy savings and the volume of drinking water production up to ~ 8.0% and ~ 9.8% respectively. The use of additional back-washing cycles will also limit the high fouling behaviour of the membrane that may occur even when the optimized back-washing as observed in this study was implemented towards the end of

the filtration. In addition, it is expected that optimal operation would further extend the life span of the membrane and minimize the need for chemical cleaning to recover flux decline caused by irreversible fouling. Current research is investigating the number of back-washing cycles as another optimization parameter.

The ability of this modeling approach to reasonably predict the reversible and irreversible fouling behaviour experienced by both membranes with back-washing time intervals that are different to those employed in the model calibration, indicate that the proposed approach is generally robust in modeling different – generally unplanned or infrequent – filtration situations which are difficult to model. However, model predictions generally appeared to slightly deviate from the experimental flux measurements towards the end of the filtration experiments. This is accepted as the model predictions are based only on the fluorescence EEMs captured at time = 15 min of the UF experiments. This inaccuracy will be addressed in future studies by using more frequent fluorescence EEM measurements for updating the model along the filtration process.

10.4.5 Role of Membrane Foulant Components in Reversible and Irreversible Fouling

Another important aspect of the modeling approach developed in this study is that it can be used to estimate the accumulation of individual foulant components in/on the membranes in terms of PC scores (i.e. $s_{j,M}$; where j is the PC related to the j^{th} foulant component) as illustrated in Equations 10.2a and 10.2b. These individual membrane foulant components contribute differently to the increase in the membrane resistance (R_t) and this relative contribution is quantified by the value of $\beta_j s_{j,M}$ for each foulant component (j) in Equation

10.3. Therefore by examining the evolution of $\beta_j s_{j,M}$ (for $j = 1, 2, 3, 4$), one can assess how different membrane fouling components, identified in this study, contribute to membrane fouling. Figure 10-4a - d illustrate the evolution of these estimates for the main foulant components, such as HS-like, protein-like and colloidal/particulate matter for high, medium and low fouling events experienced by 60 kDa membranes. Similar observations were made for 20 kDa membranes (results not shown for brevity). The corresponding PCs are also indicated in Figure 10-4.

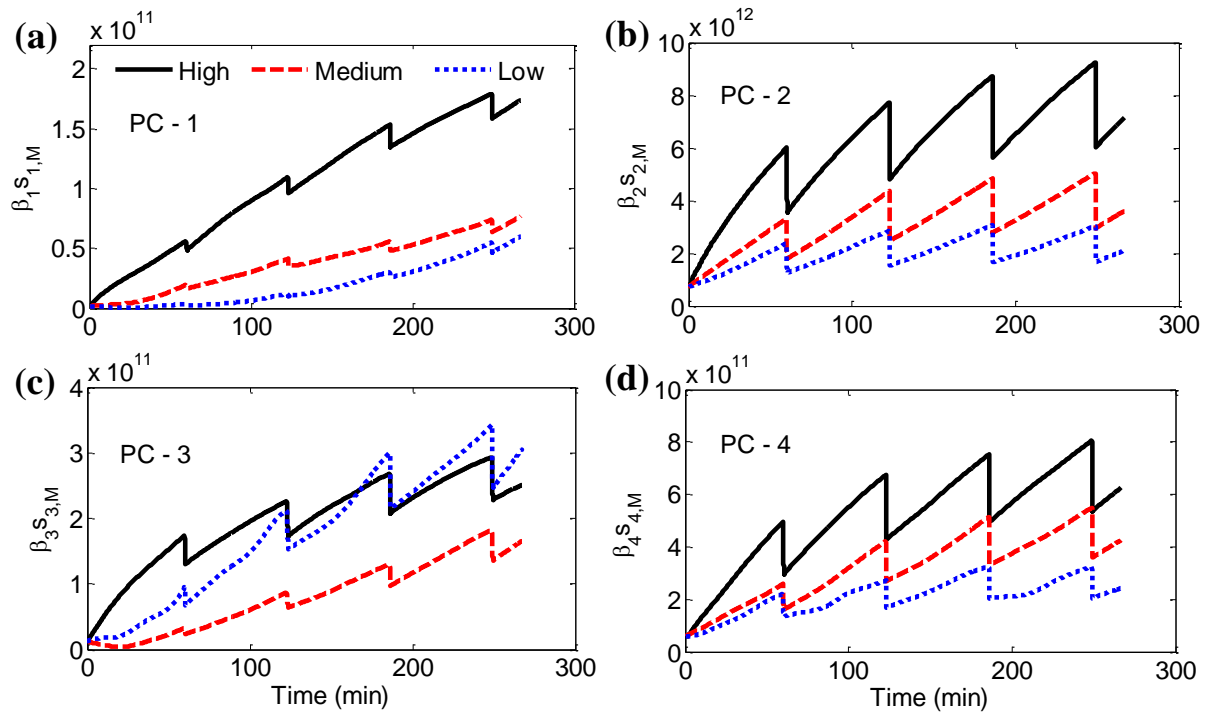


Figure 10-4 The contribution of (a) humic substances (HS) – like (PC – 1), (b and d) colloidal/particulate (PC – 2 and PC – 4) and (c) protein-like (PC – 3) matter on membrane resistance as calculated in terms of the accumulation of individual foulant components in/on the membranes for selected 60 kDa UF experiments of low, medium and high membrane fouling situations.

It should be noted that these estimates were calculated using the PC scores related to fluorescence EEMs of retentate and permeate obtained during the course of the UF

experiments using Equations 10.2a, 10.2b and 10.3. Thus, these model estimates obtained through Equation 10.2a with actual fluorescence data collected along the experiment are representative of the foulant accumulation on the membranes during the course of the 60 kDa UF experiments.

The cyclical drops in $\beta_{js_{j,M}}$ values in Figure 10-4a – d are correlated with the flux increase or alternatively fouling decrease obtained after membrane back-washing as illustrated in Figure 10-2. Based on the relatively larger changes observed after backwashing for PC – 2 and PC – 4 one can conclude that colloidal/particulate matter accumulated in/on the membranes was removed predominantly by back-washing compared to HS- and protein-like foulants. As a result, colloidal/particulate matter demonstrates a comparatively larger contribution to reversible fouling. HS- and protein-like foulants on the other hand are seen to be contributing significantly towards irreversible fouling (i.e. smaller drop in $\beta_{1s_{1,M}}$ and $\beta_{2s_{2,M}}$ values after back-washes). These observations are consistent with interpretations provided for reversible and irreversible UF fouling by river water foulant extracts (Jucker and Clark, 1994; Aoustin *et al.*, 2001) and model foulants (Jones and O'Melia, 2001; Jermann *et al.*, 2008b). Therefore, the proposed modeling approach allows identification of the type of foulant components in the water that are contributing to reversible and irreversible fouling eliminating the need to perform membrane autopsy analyses which are difficult and time consuming. In addition to drinking water treatment related membrane applications, this approach could also have application in other types of membrane-based treatment or separation of substances that have fluorescence properties.

10.5 Conclusions

The proposed fluorescence-based membrane fouling modeling approach was suitable for accurately predicting different fouling situations for UF cross-flow treatment of river water. Model forecasts/predictions are based on the fluorescence EEM measurements captured at time = 15 min of the UF operation and therefore should allow sufficient time for fouling control strategies to be implemented. The approach is especially applicable for forecasting high fouling events that are often harmful for membranes or challenging for the efficient production of drinking water to meet consumer demand. The potential of this approach for process optimization would be very useful for the sustainable operation of membrane-based drinking water treatment operations in terms of minimizing the energy spent per unit amount of drinking water produced. In addition, the proposed approach is also able to identify specific membrane foulants that contribute to reversible and irreversible fouling of membranes in drinking water applications.

10.6 Supplementary Information

10.6.1 Typical Fluorescence Spectral Features of Grand River Water

The fluorescence spectral region (α) at excitation wavelength (Ex) \sim 320 nm and emission wavelength (Em) \sim 415 nm and region (β) at Ex/Em \sim 270 nm/460 nm, indicated in the fluorescence EEM of GRW (Figure S - 10-1), are representative of the presence of HS-like NOM foulants (Coble *et al.*, 1990; Sierra *et al.*, 2005). GRW with similar fluorescence EEM properties were analyzed using liquid chromatography-organic carbon detection (LC-OCD) analysis to independently confirm the presence of HS (Peiris *et al.*, 2008 and 2010). The deviations of the fluorescence EEM contours seen in region (δ) at around Ex/Em \sim 280 nm/330 nm is related to the presence of protein-like NOM foulants in GRW (Peiris *et al.*, 2010). This observation is also supported by other studies (Baker, 2001; Chen *et al.*, 2003; Her *et al.*, 2003). Rayleigh scattering regions (RS) observed in the fluorescence EEM also provide information related to the particulate/colloidal matter present in water (Peiris *et al.*, 2010).

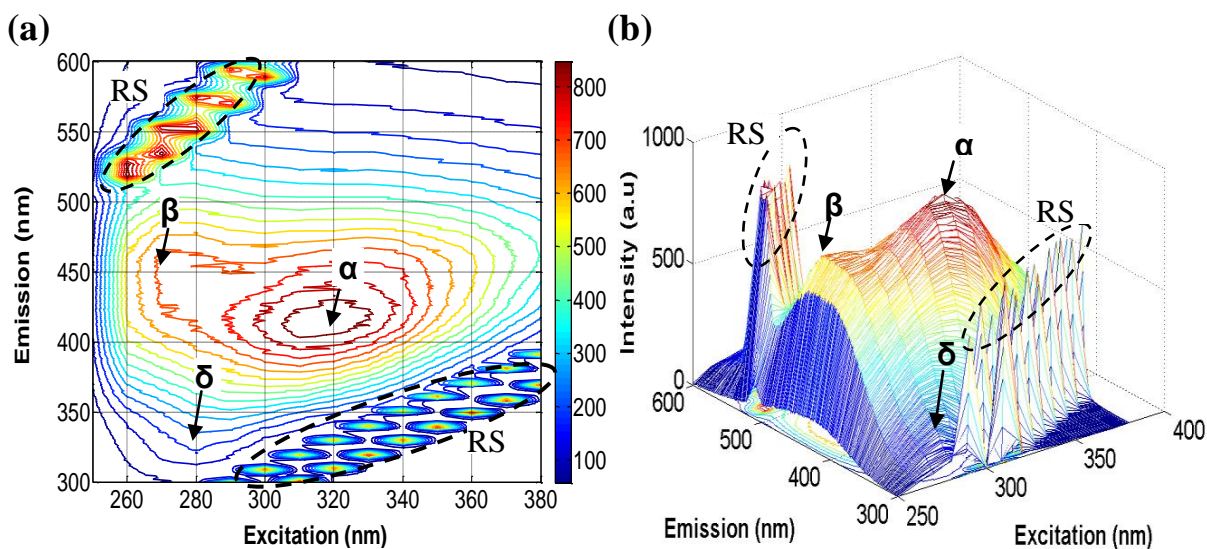


Figure S - 10-1 Typical fluorescence features seen in the (a) fluorescence EEM for GRW and (b) 3D view of the same EEM. Rayleigh light scattering (RS) regions are indicated using dashed-lines.

10.6.2 Loading Plots for 60 kDa and 20 kDa UF Membrane Experiment

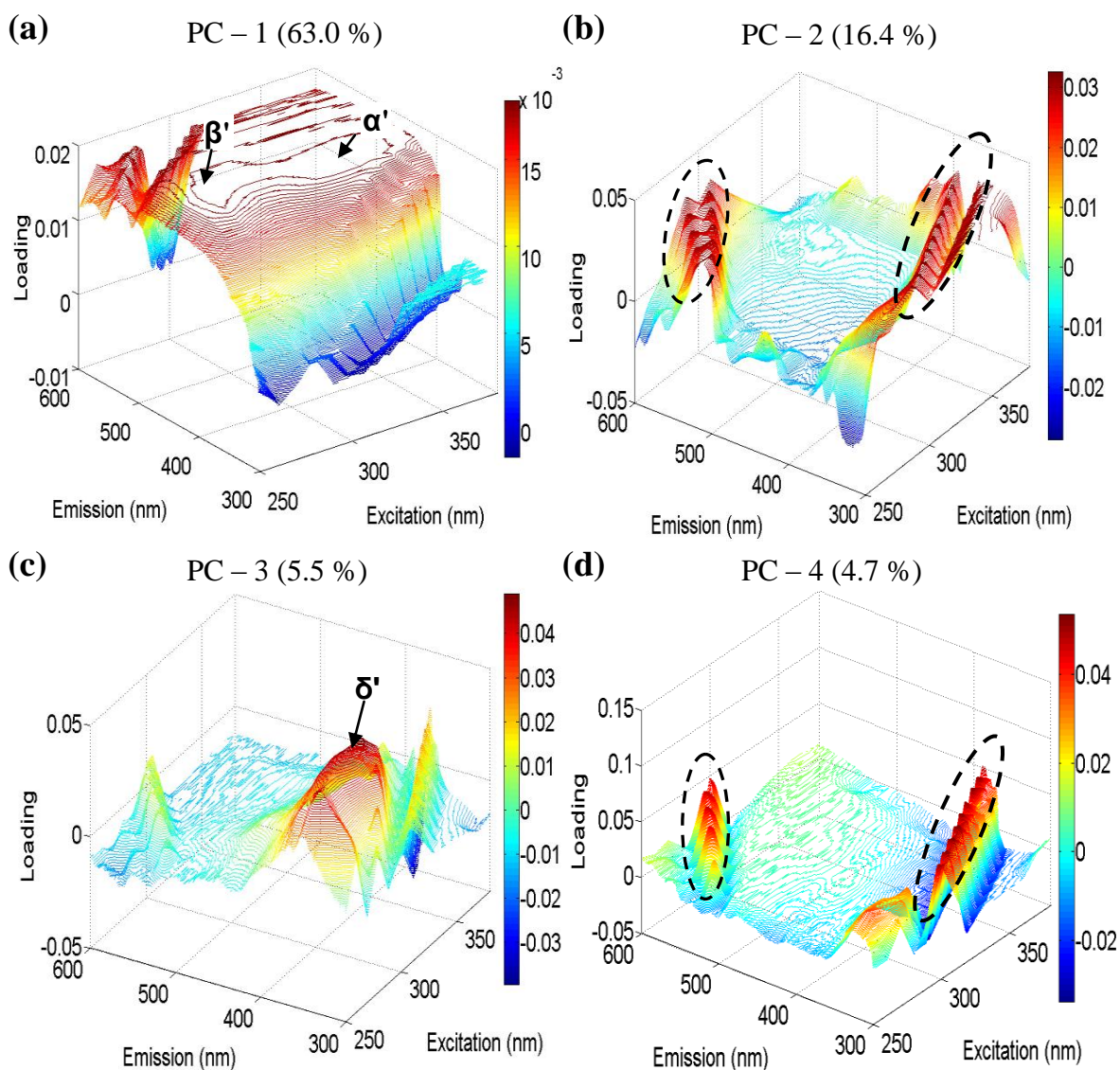


Figure S - 10-2 3D illustrations of the loading matrices of (a) PC – 1, (b) PC – 2, (c) PC – 3 and (d) PC – 4 generated by the PCA of 60 kDa data. Rayleigh scattering peak-like regions are indicated in dashed circles. Variance captured by each PC is indicated within parentheses.

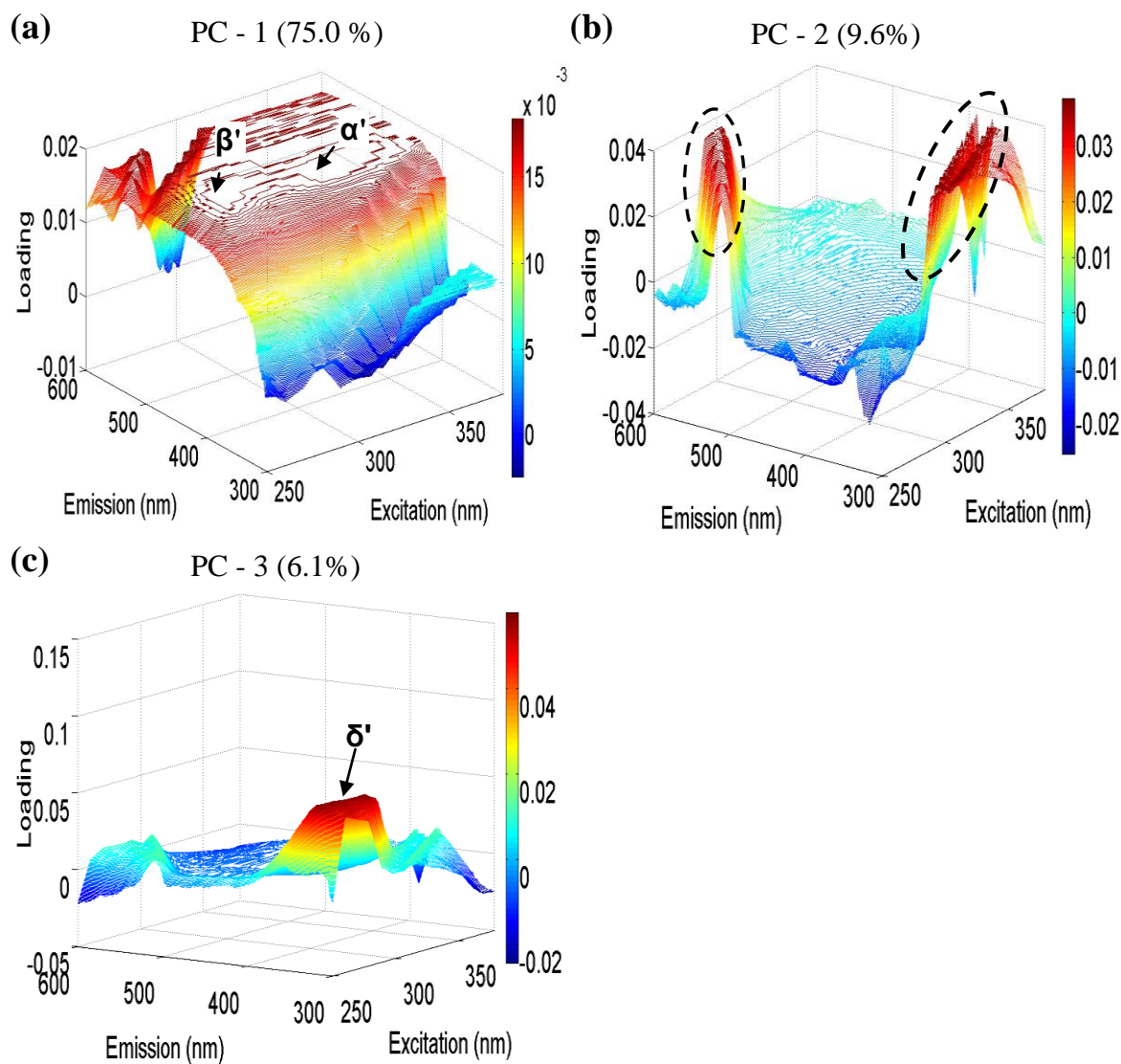


Figure S - 10-3 3D illustrations of the loading matrices of PC -1, PC-2 and PC-3 generated by the PCA of 20 kDa data. Rayleigh scattering peak-like regions are indicated in dashed circles. Variance captured by each PC is indicated within parentheses.

10.6.3 Model Parameters for Fluorescence-based Fouling Modeling Approach

Table S - 10-1 Parameters used in the modeling of 60kDa and 20kDa UF processes

Parameter	Model for 60 kDa UF membrane	Model for 20 kDa UF membrane
Calculated/measured		
A	42.09e-4 m ²	42.09e-4 ²
\dot{m}_{wash}	2.56e-7 m ³ s ⁻¹	2.56e-7 m ³ s ⁻¹
R _o	2.76e12 m ⁻¹	4.01e12 m ⁻¹
V _m	9.68e-6 m ³	9.68e-6 m ³
ΔP	103.35 kPa	103.35 kPa
M	9e-4 Nsm ⁻²	9e-4 Nsm ⁻²
Estimated*		
β ₁	1.63e7	1.44e7
β ₂	7.14e8	8.84e8
β ₃	1.18e8	1.48e8
β ₄	9.57e7	-
β _{inter}	-5.01e4	-3.96e4
eff ₁	0.1	0.2
eff ₂	0.4	0.5
eff ₃	0.2	0.3
eff ₄	0.3	-
K	0.24	0.23
q ₁	-7e-14	-1e-13
q ₂	-9e-14	-1.25e-13
q ₃	-7e-14	-1e-13
q ₄	-8e-14	-
W	1 or 0	1 or 0
Z ₁	0.67	0.60
Z ₂	0.60	0.60

* It should be noted that the units of parameters: β_j (j = 1, 2, ..., 4) and β_{inter} are not given as the scores (s_i) are considered to be unit less. Nevertheless, β_js_j and β_{inter}s_{protein}·s_{coll./partic.} have the unit “m⁻¹”

CHAPTER 11

Conclusions & Recommendations

The research presented in this thesis focused on the development of fluorescence-based methods that were found suitable for characterizing NOM and colloidal/particulate matter present in natural water. The fluorescence EEM method developed was used in combination with multivariate statistical data analyzing techniques such as PLS and PCA to develop:

1. calibration models for identification of the relative changes of major NOM fractions such as HA-, FA- protein and polysaccharide-like matter present in water.
2. monitoring tools for assessing the extent of removal of NOM and colloidal/particulate matter at different treatment stages and their contribution to membrane fouling (i.e. reversible and irreversible fouling) in drinking water treatment operations and,
3. membrane fouling models for prediction of UF membrane fouling behaviour and optimization/control strategies for membrane filtration suitable for maximizing drinking water production with minimum energy consumption.

The discussion that follows will present the main conclusions of this work accentuating the novel techniques developed as compared to the available methods reported in the literature. The advantages of these techniques and the challenges involved in implementing them will also be discussed.

11.1 Acquiring Reproducible Fluorescence Signals for Very Low NOM Concentration Levels

Acquiring reproducible fluorescence signals for very low concentration levels of NOM in water, which is typical in drinking water applications and in particular for membrane permeates, has historically been challenging. Due to this reason, most studies have used pre-concentration of water samples before fluorescence analysis. This research demonstrated that it was possible to obtain reproducible fluorescence signals with minimum signal noise by manipulating fluorescence spectrofluorometer parameters, such as PMT voltage, scanning rate and excitation/emission slit widths (Chapter 3). The approach demonstrated here was successfully applied in the analysis of very low NOM/DOM concentrations levels (similar to concentration levels of NF permeate) and could be useful for researchers interested in using fluorescence spectroscopy to analyze similar concentrations levels. Nevertheless, this approach should be practiced with care to avoid diminishing signal resolution. The advantage of this approach is that it is now possible to use fluorescence spectroscopy as an on-line or rapid off-line NOM characterization method which is essential in the process control and optimization involving drinking water treatment operations.

11.2 Fluorescence-based Qualitative Characterization of NOM

Characterization of all major NOM fractions, such as HA-, FA-, protein- and polysaccharide-like matter, present in natural water has been difficult due to their complex nature and low concentration levels. Chapter 4 highlights the difficulty in performing direct fluorescence-based analysis as NOM fractions of smaller concentrations (HA-like NOM in this case) are

overshadowed by the fluorescence signals of the more abundant FA-like NOM. A simple pH-based separation of HA- and FA-like NOM fractions was proposed followed by fluorescence EEM analysis that could be helpful for obtaining more accurate qualitative information on the HA- and FA-like NOM content in natural water; however, this separation method is not capable of separating other NOM fractions including protein- and polysaccharide-like matter.

11.3 Fluorescence-based Calibration Models for Characterization of NOM

Many researchers have used fluorescence spectroscopy as a tool for NOM characterization, but absolute quantification methods have neither been reported nor has any form of fluorescence-based characterization of polysaccharide-like NOM fractions been possible. The proposed fluorescence-based calibration modeling approach can be used to perform rapid, simultaneous and direct characterization of all the four major NOM fractions, namely, HA-, FA-, protein- and polysaccharide-like matter. This method is suitable for identifying the relative changes of NOM fractions present in water in terms of DOC with good accuracy within the concentration levels that are typical to drinking water applications. Compared to other available fluorescence-based NOM characterization methods, this approach showed better accuracy and less sensitivity to noise. In addition, the ability to rapidly characterize polysaccharide-like NOM and interactions between polysaccharide- and protein-like NOM is novel and possibly of considerable practical value. At high NOM concentrations (i.e. outside of calibration ranges), non-linear correlations present in the fluorescence data could affect the accuracy of the estimations.

11.4 Process Monitoring in Drinking Water Treatment Systems using Fluorescence EEM Measurements

PCA of fluorescence EEMs was identified as a promising approach to quantify the impact of pre-treatment stages on the removal of HS, protein- and colloidal/particulate matter foulants for UF/NF membranes (Chapter 8). PC scores, generated with this approach, were shown to be related to these membrane foulants. The higher sensitivity of this approach, compared to more conventional process monitoring parameters like TOC and turbidity measurements used for drinking water treatment systems, was demonstrated to be suitable for identifying high membrane fouling events that occur due to inadequate treatment in the pre-treatment stages of these membrane-based treatment processes. The proposed approach has the potential for use as a monitoring tool for membrane-based water treatment and pre-treatment operations, and as an early detection method to identify high fouling events. This would allow membrane operational changes to be implemented proactively.

This method was performed using fluorescence EEM measurements which required ~ 5 min for each water sample analyzed. Therefore the approach is suitable for process control situations where a small time delay (~ 5 min) between operational changes and fluorescence measurements is acceptable. Since the time frames involved with membrane fouling in drinking water treatment applications would normally be in the order of hours or more, this approach could be used for near real-time or rapid off-line monitoring.

11.5 Understanding Membrane Fouling Behaviour

The feasibility of using fluorescence EEMs of NF permeates to assess the fouling behaviour of two types of NF membranes was demonstrated. Although TOC, DOC, turbidity and LC-OCD chromatographic measurements offered little help and were not sensitive enough to identify the different rejection characteristics that contributed to different fouling behaviour experienced by the two NF membranes, the differences captured by fluorescence EEMs appeared to have a correlation with the fouling behaviour observed. This ability can be attributed to the higher sensitivity that is available in the fluorescence measurement procedure developed in this thesis (Chapter 3).

Fluorescence EEMs of membrane permeates are not sufficient for achieving a more comprehensive understanding of the membrane fouling behaviour involving hydraulically reversible and irreversible membrane fouling. This was successfully addressed by a novel approach that combined PCA analysis and fluorescence EEM measurements to characterize three major membrane foulant fractions: HS, protein- and colloidal/particulate. Each of these components was related to specific PCs generated by the PCA of fluorescence EEMs captured during the course of UF. Using this approach the accumulation of these major foulant components on the membrane could be monitored by examining the temporal evolution of the PC score difference between the retentate and permeate during UF of representative river water (GRW). In particular, it was demonstrated that colloidal/particulate components contributed mainly to reversible fouling of UF membranes, while HS- and protein-like components contributed more significantly to irreversible fouling.

Much of the research work reported in literature has focused on using membrane autopsy analyses and filtration studies using model membrane foulants to understand the contribution of different foulant fractions in water for reversible and irreversible fouling. In this context, this novel approach presents a significant potential as a fouling monitoring tool in the field and in membrane research. By employing this method, it is possible to understand the fouling behaviour of natural water during filtration operations without having to perform membrane autopsy analyses.

11.6 Membrane Fouling Modeling and Optimization

The lack of understanding of different physicochemical phenomena involved in membrane fouling has impeded the development of mechanistic approaches for modeling membrane fouling in drinking water applications. With the fluorescence-based membrane foulant characterization methods developed in this research to characterize the drinking water membrane foulants, it was possible to understand how these membrane foulants contributed to different UF membrane fouling behaviour involving reversible and irreversible fouling. This improved understanding enabled the development of a fluorescence-based membrane fouling modeling approach that was suitable for accurately predicting different fouling situations. The approach was able to forecast high fouling events that are generally harmful for membranes or challenging for the efficient production of drinking water to meet consumer demand. The feasibility of using this approach for optimizing UF of drinking water was also demonstrated. The potential of this approach for process optimization would be very useful for the sustainable operation of membrane-based drinking water treatment

facilities in terms of minimizing the energy consumed per unit amount of drinking water produced.

11.7 Future Research Work

1. Fluorescence-based calibration approaches developed for characterization of NOM (Chapters 5 and 6) need to be validated with independent measurement techniques that are able to differentiate different NOM fractions from natural water samples. In the absence of such techniques that are able to characterize all major NOM fractions, independent validation of this technique will be a challenge. The LC-OCD chromatographic method however, is able to characterize NOM in terms of HS and biopolymers (polysaccharide and protein in combination), but not in terms of HA-, FA-, protein and polysaccharide-like matter. Therefore it is proposed that future research should be directed towards using the fluorescence-based calibration and LC-OCD techniques to find correlations between the calibration model predictions and the concentration of NOM fractions that could be estimated by the LC-OCD method for natural water. This will enable an indirect validation of this approach. In addition, characterization of natural water samples from different water sources can provide an understanding of the limitations of this approach and identify areas for future improvements.
2. It is also expected that the NOM characterization techniques described in this thesis could be used to estimate the disinfection by-products formation potential in water.

Research work focusing on this area would be of significant benefit for the drinking water treatment industry.

3. The fluorescence-based process monitoring techniques (Chapters 7, 8 and 9) and the fouling modeling and optimization approach (Chapter 10) discussed in this thesis were developed using lab/bench-scale treatment systems. Therefore, these techniques need to be tested to confirm their feasibility for monitoring of membrane fouling and treatment processes in large-scale drinking water treatment facilities that experience different sources of raw water containing varying degrees of NOM and colloidal/particulate matter content.
4. The experimental work involved in the development of the fluorescence-based fouling modeling and optimization approach was conducted using flat sheet membrane systems. Future research work directed at using this technique on MF, UF, NF and RO membranes with different configurations such as spiral wound, hollow fibre and tubular membranes would provide a better understanding of the limitation of this approach and areas for improvement.
5. The fluorescence-based techniques developed in this research were based on cuvette-based analysis. With this mode of analyses, only a limited number of samples could be analyzed at a given time. Analyses of a large number of water samples could be possible by using a fluorescence plate reader. However, future research that focuses on optimization of instrumental parameters (similar to work presented in Chapter 3) should be conducted to ensure the acquisition of

reproducible fluorescence signals using this method. In addition, development of fluorescence-based approaches suitable for on-line or rapid off-line analyses with a shorter signal acquisition time is also an important area for future research.

REFERENCES

- Aiken, G.R. 1985. Isolation and concentration techniques for aquatic humic substances. In: Aiken, G.R., McKnight, D.M., Wershaw, R.L., MacCarthy, P. (Eds.), *Humic Substances in Soil, Sediment, and Water*, Wiley, New York., 363-385 pp.
- Alberts, J.J., Takács, M. 2004. Total luminescence spectra of IHSS standard and reference fulvic acids, humic acids and natural organic matter: comparison of aquatic and terrestrial source terms. *Org. Geochem.* **35** (3), 243-256.
- Alberts, J.J., Takács, M., McElvaine, M., Judge, K. 2001. Apparent size distribution and spectral properties of natural organic matter isolated from six rivers in southeastern Georgia, USA. In: Ghabbour, E., Davies, G. (Eds.), *Humic Substances, Structures, Models and Functions*, The Royal Society of Chemistry, Cambridge, UK, 179-190 pp.
- Amy, G. 2008. Fundamental understanding of organic matter fouling of membranes. *Desalination*, **231**, 44-51.
- Aoustin, E., Schäfer, A.I., Fane, A.G., Waite, T.D. 2001. Ultrafiltration of natural organic matter, *Sep. Purif. Technol.*, **22–23** (1–3) 63–78.
- Atkinson, S. 2004. UF membranes produce drinking water for US communities. *Membr. Technol.* **10**, 10-16.
- Baker, A. 2001. Fluorescence excitation - Emission matrix characterization of some sewage-impacted rivers. *Environ. Sci. and Technol.* **35** (5), 948–953.

- Boehl, D., Solle, D., Hitzmann, B., Scheper, T. 2003. Chemometric modelling with two-dimensional fluorescence data for *Claviceps purpurea* bioprocess characterization. J. Biotechnol. **105** (1-2), 179-188.
- Boehme, J., Coble, P., Conmy, R., Stovall-Leonard, A. 2004. Examining CDOM fluorescence variability using principal component analysis: seasonal and regional modeling of three dimensional fluorescence in the Gulf of Mexico. Mar. Chem. **89** (1-4), 3–14.
- Bolton, G., LaCasse, D., Kuriyel, R. 2006. Combined models of membrane fouling: Development and application to microfiltration and ultrafiltration of biological fluids. J. Membr. Sci. **277** (1-2), 75-84.
- Bowen, W.R., Calvo, J.I., Hernandez, A. 1995. Steps of membrane blocking in flux decline during protein microfiltration. J. Membr. Sci. **101** (1-2), 153.
- Braghetta, A., DiGiano, F.A., Ball, W.P. 1997. Nanofiltration of natural organic matter: pH and ionic strength effects. J. Environ. Eng. **123** (7), 628-641.
- Braghetta, A., DiGiano, F.A., Ball, W.P. 1998. NOM accumulation at NF membrane surface: Impact of chemistry and shear. J. Environ. Eng. **124** (11), 1087-1097.
- Cabassud, M., Delgrange-Vincent, N., Cabassud, C., Durand-Bourlier, L., Laine, J.M. 2002. Neural networks: A tool to improve UF plant productivity. Desalination **145** (1), 223-231.

- Carroll, T., King, S., Gray, S.R., Bolto, B.A., Booker, N.A. 2000. Fouling of microfiltration membranes by NOM after coagulation treatment. *Water Res.* **34** (11), 2861-2868.
- Casado-Terrones, S., Fernández-Sánchez, J., Segura-Carretero, A., Fernández-Gutiérrez, A. 2007. Simple luminescence detectors using a light-emitting diode or a Xe lamp, optical fiber and charge-coupled device, or photomultiplier for determining proteins in capillary electrophoresis: A critical comparison. *Anal. Biochem.* **365** (1), 82-90.
- Chang, Y., Benjamin, M.M. 2003. Modeling formation of natural organic matter fouling layers on ultrafiltration membranes. *J. Environ. Eng.* **129** (1), 25-32.
- Chen, J., LeBoeuf, E.J., Dai, S., Gu, B. 2003. Fluorescence spectroscopic studies of natural organic matter fractions. *Chemosphere* **50** (5), 639-647.
- Chen, W., Westerhoff, P., Leenheer, J.A., Booksh, K. 2003. Fluorescence excitation-emission matrix regional integration to quantify spectra for dissolved organic matter. *Environ. Sci. and Technol.* **37** (24), 5701–5710.
- Cho, J., Amy, G., Pellegrino, J. 1999. Membrane filtration of natural organic matter: initial comparison of rejection and flux decline characteristics with ultrafiltration and nanofiltration membranes. *Water Res.* **33** (11), 2517-2526.
- Cho, J., Amy, G., Pellegrino, J. 2000. Membrane filtration of natural organic matter: Factors and mechanisms affecting rejection and flux decline with charged ultrafiltration (UF) membrane. *J. Membr. Sci.* **164** (1-2), 89-110.

- Christman, R.F., Norwood, D.L., Seo, Y., Frimmel, F.H. 1989. Oxidative degradation of humic substances from freshwater environments. In: Hayes, M.H.B., MacCarthy, P., Malcolm, R.L., Swift, R.S. (Eds.), Humic substances; II, In search of structure, John Wiley & Sons, Chichester, United Kingdom 34–67 pp.
- Christy, A.A., Egeberg, P.K. 2000. Characterisation of natural organic matter from the Nordic typing project water samples by chemometric analysis of their near infrared spectral profiles. *Chemo. and Intell. Lab. Sys.* **50** (2), 225-234.
- Clark, M.M., Lucas, P. 1998. Diffusion and partitioning of humic acid in a porous ultrafiltration membrane. *J. Membr. Sci.* **143** (1-2), 13-25.
- Coble, P.G. 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar. Chem.* **51** (4), 325-346.
- Coble, P.G., Green, S.A., Blough, N.V., Gagosian, R.B. 1990. Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. *Nature* **348** (6300), 432-435.
- Combe, C., Molis, E., Lucas, P., Riley, R., Clark, M.M. 1999. Effect of CA membrane properties on adsorptive fouling by humic acid. *J. Membr. Sci.* **154** (1), 73-87.
- Creighton, T.E. 1997. Protein Structure - A Practical Approach. In: Hames, B.D. (Ed.), The Practical Approach Series, Oxford University Press Inc., New York, 270-283 pp.
- Croué, J. 2004. Isolation of humic and non-humic NOM fractions: Structural characterizations. *Environ. Monit. Assess.* **92** (1-3), 193-207.

- Croué, J. P., Korshin, G. V., Leenheer, J., Benjamin, M.M. 2000. Isolation, fractionation and characterization of natural organic matter in drinking water. AWWA Research Foundation and American Water Works Association Denver, Colo. 324 pp.
- Del Castillo, Carlos E., Coble, P. G., Morell, J.M., López, J.M., Corredor, J.E. 1999 Analysis of the optical properties of the Orinoco River plume by absorption and fluorescence spectroscopy. *Mar. Chem.* **66** (1), 35-51.
- Delgrange-Vincent, N., Cabassud, C., Cabassud, M., Durand-Bourlier, L., Laine, J.M. 2000. Neural networks for long term prediction of fouling and backwash efficiency in ultrafiltration for drinking water production. *Desalination* **131** (1), 353-362.
- Ebrahim, S. 1994. Cleaning and regeneration of membranes in desalination and wastewater applications: state-of-the-art. *Desalination* **96** (1-3), 225-238.
- Elshereef, R., Budman, H., Moresoli, C., Legge, R.L. 2006. Fluorescence spectroscopy as a tool for monitoring solubility and aggregation behavior of alpha & beta-lactoglobulin after heat treatment. *Biotechnol. Bioeng.* **95** (5), 863-874.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S. 2001. Multi- and Megavariate Data Analysis, Principles and Applications. Umetrics Academy, Umea, Sweden, ISBN 91-973730-1-X, p. 533.
- Esbensen, K.H. 2002. Multivariate data analysis - in practice. Camo Process AS, Oslo N-0158, Norway, p. 598.

- Fabris, R., Lee, E.K., Chow, C.W.K., Chen, V., Drikas, M. 2007. Pre-treatments to reduce fouling of low pressure micro-filtration (MF) membranes. *J. Membr. Sci.* **289** (1-2), 231-240.
- Fan, L., Harris, J.L., Roddick, F.A., Booker, N.A. 2001. Influence of the characteristics of natural organic matter on the fouling of microfiltration membranes. *Water Res.* **35** (18), 4455-4463.
- Field, A. 2000. *Discovering Statistics Using SPSS for Windows*. Sage Publications, London, 496 pp.
- Fiksdal, L., Leiknes, T. 2006. The effect of coagulation with MF/UF membrane filtration for the removal of virus in drinking water. *J. Membr. Sci.* **279** (1-2), 364-371.
- Freeman, S., William, V. 2005. Operations. In: Christensen, M. (Ed.), *Microfiltration and Ultrafiltration Membranes for drinking Water- 1st ed.* (AWWA manual: M53), American Water Works Association, Denver Co. 187-203 pp.
- Fu, P., Wu, F., Liu, C. 2004. Fluorescence excitation-emission matrix characterization of a commercial humic acid. *Chinese J. Geochem.* **23** (4), 309-318.
- Ghabbour, E.A., Khairy, A.H., Cheney, D.P., Gross, V., Davies, G., Gilbert, T.R., Zhang, X. 1994. Isolation of humic acid from the brown alga *Pilayella littoralis*. *J. Appl. Phycol.* **6** (5), 459-468.
- Glucina, K., Alvarez, A., Laîné, J.M. 2000. Assessment of an integrated membrane system for surface water treatment. *Desalination* **132** (1-3), 73-82.

- Gray, S.R., Ritchie, C.B., Tran, T., Bolto, B.A. 2007. Effect of NOM characteristics and membrane type on microfiltration performance. *Water Res.* **41** (17), 3833-3841.
- Gunst, F.R., Mason, R.L., Hua, T.A. 1981. Classroom Supplement to Regression Analysis and Its Application: A Data Oriented Approach. In: Anonymous Marcel Dekker Inc. Newyork and Basel, 29-30 pp.
- Haberkamp, J. 2008. Organisches Membranfouling bei der Ultrafiltration kommunaler Kläranlagenabläufe-Ursachen, Mechanismen und Maßnahmen zur Verringerung, Ph.D Dissertation, Technische Universität Berlin, Berlin, Germany.
http://opus.kobv.de/tuberlin/volltexte/2009/2107/pdf/haberkamp_jens.pdf
- Hallé, C., Huck, P.M., Peldszus, S., Haberkamp, J., Jekel, M. 2009. Assessing the performance of biological filtration as pretreatment to low pressure membranes for drinking water. *Environ. Sci. and Technol.* **43** (10), 3878-3884.
- Hamed, O.A. 2005. Overview of hybrid desalination systems - Current status and future prospects. *Desalination* **186** (1-3), 207-214.
- Henderson, R.K., Baker, A., Murphy, K.R., Hambly, A., Stuetz, R.M., Khan, S.J. 2009. Fluorescence as a potential monitoring tool for recycled water systems: A review. *Water Res.* **43** (4), 863-881.
- Her, N., Amy, G., McKnight, D., Sohn, J., Yoon, Y. 2003. Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. *Water Res.* **37** (17), 4295–4303.

- Her, N., Amy, G., Plottu-Pecheux, A., Yoon, Y. 2007. Identification of nanofiltration membrane foulants. *Water Res.* **41** (17), 3936-3947.
- Her, N., Amy, G., Sohn, J., Gunten, U. 2008. UV absorbance ratio index with size exclusion chromatography (URI-SEC) as an NOM property indicator. *J. Water Supply: Res. and Technol.-AQUA.* **57** (1), 35-44.
- Herman, B., Lakowicz, J.R., Murphy, D.B., Spring, K.R., Davidson, M.W. 2003. Fluorescence Microscopy - Basic Concepts in Fluorescence. cited on April 4, 2008 from <http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorescenceintro.html>
- Holzalski, R.M., Goel, S., Bouwer, E.J. 1995. TOC removal in biological filters, *JAWWA*, **87** (12), 40-54.
- Hong, S., Elimelech, M. 1997. Chemical and physical aspects of natural organic matter (NOM) fouling of nanofiltration membranes. *J. Membr. Sci.* **132** (2), 159-181.
- Huber, S.A., Frimmel, F.H. 1992. A liquid chromatographic system with multi-detection for the direct analysis of hydrophilic organic compounds in natural waters. *Fresenius J. Anal. Chem.* **342** (1-2), 198-200.
- Huck, P.M. 1999. Development of a framework for quantifying the removal of humic substances by biological filtration. *Water Sci. and Technol.* **40** (9), 149-156.
- Hudson, N., Baker, A., Reynolds, D. 2007. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters - A review. *River Res. and Appl.* **23** (6), 631-649.

- Hudson, N., Baker, A., Ward, D., Reynolds, D.M., Brunsdon, C., Carliell-Marquet, C., Browning, S. 2008. Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water quality assessment? An example from South West England. *Sci. Total Environ.* **391** (1), 149-158.
- Imeson, A.P., Ledward, D.A., Mitchell, J.R. 1997. On the nature of the interaction between some anionic polysaccharides of pectate and alginate gels. *J. Sci. Food Agric.* **28** (8), 661-668.
- Jeanneau, L., Faure, P., Jarde, E., 2007. Influence of natural organic matter on the solid-phase extraction of organic micropollutants. Application to the water-extract from highly contaminated river sediment. *J. Chromatogr. A* **1173** (1-2), 1-9.
- Jermann, D., Pronk, W., Boller, M. 2008a. Mutual influences between NOM and inorganic particles and their combined effect on UF membrane fouling. *Environ. Sci. and Technol.* **42** (24), 9129-9136.
- Jermann, D., Pronk, W., Boller, M., Schafer, A.I., 2009. The role of NOM fouling for the retention of estradiol and ibuprofen during ultrafiltration. *J. Membr. Sci.* **329** (1-2), 75-84.
- Jermann, D., Pronk, W., Kagi, R., Halbeisen, M., Boller, M. 2008b. Influence of interactions between NOM and particles on UF fouling mechanisms. *Water Res.* **42** (14), 3870-3878.
- Jermann, D., Pronk, W., Meylan, S., Boller, M. 2007. Interplay of different NOM fouling mechanisms during ultrafiltration for drinking water production. *Water Res.* **41** (8), 1713–1722.

- Jiang, T. 2007. Characterization and modelling of soluble microbial products in membrane bioreactors, Ph.D Dissertation, Ghent University, Belgium. pp. 14.
www.biomath.ugent.be/publications/download/jiangtao_sum.pdf
- Jones, K. L., O'Melia, C. R. 2001. Ultrafiltration of protein and humic substances: effect of solution chemistry on fouling and flux decline. *J. Membr. Sci.* **193** (2), 163-173.
- Jones, K.L., O'Melia, C.R. 2000. Protein and humic acid adsorption onto hydrophilic membrane surfaces: Effects of pH and ionic strength. *J. Membr. Sci.* **165** (1), 31-46.
- Jucker, C., Clark, M.M. 1994. Adsorption of aquatic humic substances on hydrophobic ultrafiltration membranes. *J. Membr. Sci.* **97**, 37-52.
- Kaiya, Y., Itoh, Y., Takizawa, S., Fujita, K., Tagawa, T. 2000. Analysis of organic matter causing membrane fouling in drinking water treatment. *Water Sci. and Technol.* **41** (10), 59-67.
- Kilduff, J.E., Mattaraj, S., Pieracci, J.P., Belfort, G. 2000. Photochemical modification of poly(ether sulfone) and sulfonated poly(sulfone) nanofiltration membranes for control of fouling by natural organic matter. *Desalination* **132** (1-3), 133-142.
- Kimura, K., Hane, Y., Watanabe, Y. 2005. Effect of pre-coagulation on mitigating irreversible fouling during ultrafiltration of a surface water. *Water Sci. and Technol.* **51** (6-7), 93-100.
- Kimura, K., Yamamura, H., Watanabe, Y. 2006. Irreversible fouling in MF/UF membranes caused by Natural Organic Matters (NOMs) isolated from different origins. *Sep. Sci. Technol.* **41** (7), 1331-1344.

- Konieczny, K. 2002. Modelling of membrane filtration of natural water for potable purposes. *Desalination* **143** (2), 123-139.
- Kulovaara, M., Metsamuuronen, S., Nystrom, M. 1999. Effects of aquatic humic substances on a hydrophobic ultrafiltration membrane. *Chemosphere* **38** (15), 3485-3496.
- Kwaambwa, H.M., Maikokera, R. 2007. A fluorescence spectroscopic study of a coagulating protein extracted from *Moringa oleifera* seeds. *Coll. and Surf. B: Biointerf.* **60** (2), 213-20.
- Kweon, J.H., Lawler, D.F. 2005. Investigation of membrane fouling in ultrafiltration using model organic compounds. *Water Sci. and Technol.* **51** (6-7), 101-106.
- Lakowicz, J.R. 2006. *Principle of Fluorescence Spectroscopy*. Springer Science+Business Media, NY.
- Lee, N., Amy, G., Croué, J. 2006. Low-pressure membrane (MF/UF) fouling associated with allochthonous versus autochthonous natural organic matter. *Water Res.* **40** (12), 2357-2368.
- Lee, T., Oh, H., Choung, Y., Oh, S., Jeon, M., Kim, J.H., Nam, S.H., Lee, S. 2009. Prediction of membrane fouling in the pilot-scale microfiltration system using genetic programming. *Desalination* **247** (1-3), 285-294.
- Lewzey, J. 2007. Varian Canada Inc. Personal communication.

- Li, C., Benjamin, M.M., Korshin, G.V. 2006. Characterization of NOM and its adsorption by iron oxide coated sand (IOCS) using UV and fluorescence spectroscopy. *J. Environ. Eng. and Sci.* **5** (6), 467-472.
- Lin, C., Lin, T., Hao, O.J. 2000. Effects of humic substance characteristics on UF performance. *Water Res.* **34** (4), 1097-1106.
- Malcolm, R.L., MacCarthy, P., 1986. Limitations in the use of commercial humic acids in water and soil research. *Environ. Sci. and Technol.* **20**, 904-911.
- Marhaba, T.F., Kochar, I.H. 2000. Rapid prediction of disinfection by-product formation potential by fluorescence. *Environ. Eng. Policy* **2** (1), 29-36.
- Marhaba, T.F., Washingtton, M.B. 1998. Drinking water disinfection and by-products: history and current practice. *Adv. Environ. Res.* **1**, 103-115.
- Matthews, B.J.H., Jones, A.C., Theodorou, N.K., Tudhope, A.W. 1996. Excitation-emission-matrix fluorescence spectroscopy applied to humic acid bands in coral reefs. *Mar. Chem.* **55** (3), 317-332.
- Mobed, J.J., Hemmingsen, S.L., Autry, J.L., McGown, L.B. 1996. Fluorescence characterization of IHSS humic substances: total luminescence spectra with absorbance correction. *Environ. Sci. Technol.* **30** (10), 3061-3066.
- Mopper, K., Schultz, C.A. 1993. Fluorescence as a possible tool for studying the nature and water column distribution of DOC components. *Mar. Chem.* **41** (1-3), 229-238.

- Mounier, S., Patel, N., Quilici, L., Benaim, J.Y., Benamou, C. 1999. Fluorescence 3D de la matiere organique dissoute du Fleuve Amazone; Three dimensional fluorescence of the dissolved organic carbon in the Amazon River. *Water Res.* **33** (6), 1523-1533.
- Nilson, J.A., DiGiano, F.A. 1996. Influence of NOM composition on nanofiltration. *J. Am. Water Works Assoc.* **88** (May), 53-66.
- Parlanti, E., Wörz, K., Geoffroy, L., Lamotte, M. 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Org. Geochem.* **31** (12), 1765-1781.
- Peiris B.R.H., Budman, H., Moresoli, C., Legge, R.L. 2009. Acquiring reproducible fluorescence spectra of dissolved organic matter at very low concentrations. *Wat. Sci. And Technol.* **60** (6), 1385-1392.
- Peiris, B.R.H., Hallé, C., Haberkamp, J., Legge, R.L., Peldszus, S., Moresoli, C., Budman, H., Amy, G., Jekel, M., Huck, P.M. 2008. Assessing nanofiltration fouling in drinking water treatment using fluorescence fingerprinting and LC-OCD analyses. *Water Sci. and Technol.: Water Supply* **8** (4), 459-465.
- Peiris, R.H., Hallé, C., Budman, H., Moresoli, C., Peldszus, S., Huck, P.M., Legge, R.L. 2010. Identifying fouling events in a membrane-based drinking water treatment process using principal component analysis of fluorescence excitation-emission matrices. *Water Res.* **44** (1), 185-194.
- Persson, T., Wedborg, M. 2001. Multivariate evaluation of the fluorescence of aquatic organic matter. *Anal. Chim. Acta* **434**, 179–192.

- Peuravuori, J., Pihlaja, K. 1997. Molecular size distribution and spectroscopic properties of aquatic humic substances. *Anal. Chim. Acta* **337** (2), 133.
- Qin, S.J., McAvoy, T.J. 1992. Nonlinear PLS modeling using neural networks. *Comput. Chem. Eng.* **16** (4), 379-391.
- Radjenović, J., Petrović, M., Ventura, F., Barceló, D. 2008. Rejection of pharmaceuticals in nanofiltration and reverse osmosis membrane drinking water treatment. *Water Res.* **42** (14), 3601-3610.
- Ring, E., Samblebe, M., Leak, M., Gray, S. 2004. NOM removal for extension of chlorine dioxide residuals and lower biological regrowth potentials. *Water Sci. and Technol.: Water Supply* **4** (4), 251-254.
- Rinnan, A., Andersen, C.M. 2005. Handling of first-order Rayleigh scatter in PARAFAC modelling of fluorescence excitation-emission data. *Chemo. Intell. Lab. Syst.* **76** (1), 91-9.
- Rosipal, R., Kramer, N. 2006. Overview and recent advances in partial least squares. In: *Subspace, Latent Structure and Feature Selection - Statistical and Optimization Perspectives Workshop, SLSFS 2005*, 34-51.
- Sahoo, G.B., Ray, C. 2006. Predicting flux decline in crossflow membranes using artificial neural networks and genetic algorithms, *J. Membr. Sci.*, **283**, 147–157.

- Saravia, F., Zwiener, C., Frimmel, F.H. 2006. Interactions between membrane surface, dissolved organic substances and ions in submerged membrane filtration. *Desalination* **192** (1-3), 280–287.
- Seidel, A., Elimelech, M. 2002. Coupling between chemical and physical interactions in natural organic matter (NOM) fouling of nanofiltration membranes: Implications for fouling control. *J. Mem. Sci.* **203**, 245–255.
- Sharpless, C.M., McGown, L.B. 1999. Effects of Aluminum-Induced Aggregation on the Fluorescence of Humic Substances. *Environ. Sci. Technol.* **33** (18), 3264-3270.
- Shengji, X., Juanjuan, Y., Naiyun, G. 2008. An empirical model for membrane flux prediction in ultrafiltration of surface water. *Desalination* **221** (1-3), 370-375.
- Sierra, M.M.D., Giovanela, M., Parlanti, E., Soriano-Sierra, E.J. 2005. Fluorescence fingerprint of fulvic and humic acids from varied origins as viewed by single-scan and excitation emission matrix techniques. *Chemosphere* **58** (6), 715–733.
- Smeulders, D.E., Wilson, M.A., Patney, H., Armstrong, L. 2000. Structure of molecular weight fractions of Bayer humic substances. 2. Pyrolysis behavior of high-temperature products. *Ind. Eng. Chem. Res.* **39** (10), 3631-3639.
- Smith, P.J., Vigneswaran, S., Ngo, H.H., Ben-Aim, R., Nguyen, H. 2005. Design of a generic control system for optimising back flush durations in a submerged membrane hybrid reactor. *J. Membr. Sci.* **255** (1-2), 99-106.

- Smith, P.J., Vigneswaran, S., Ngo, H.H., Ben-Aim, R., Nguyen, H. 2006. A new approach to backwash initiation in membrane systems. *J. Membr. Sci.* **278** (1-2), 381-389.
- Spencer, R.G.M., Bolton, L., Baker, A. 2007. Freeze/thaw and pH effects on freshwater dissolved organic matter fluorescence and absorbance properties from a number of UK locations. *Water Res.* **41** (13), 2941-2950.
- Standard Methods for the Examination of Water and Wastewater. 2005. 21st ed., American Public Health/American Water Works Association/Water Environment Federation, Washington DC, USA.
- Stedmon, C.A., Markager, S., Bro, R. 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* **82**, 239–254.
- Stramski, D., Wozniak, S.B., 2005. On the role of colloidal particles in light scattering in the ocean. *Limnol. Oceanogr.* **50** (5), 1581–1591.
- Susanto, H., Arafat, H., Janssen, E. M. L., Ulbricht, M. 2008. Ultrafiltration of polysaccharide–protein mixtures: Elucidation of fouling mechanisms and fouling control by membrane surface modification. *Sep. and Purif. Technol.* **63** (3), 558-565.
- Taniguchi, M., Kilduff, J.E., Belfort, G. 2003. Modes of Natural Organic Matter Fouling during Ultrafiltration. *Environ. Sci. Technol.* **37** (8), 1676-1683.
- Tansel, B., Bao, W.Y., Tansel, I.N. 2000. Characterization of fouling kinetics in ultrafiltration systems by resistances in series model. *Desalination*, **129** (1), 7-14.

- Ter Braak, C.J.F., Verdonschot, P.F.M. 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquat. Sci.* **57** (3), 255-264.
- Thorsen, T. 1999. Membrane filtration of humic substances - state of the art. *Water Sci. and Technol.* **40** (9), 105-112.
- Thurman, E.M. 1985. *Organic Geochemistry of Natural Waters*. Martinus Nijhoff/Dr W. Junk, Dordrecht, The Netherlands, 29-41, 104-105, 282-283, 355 pp.
- Trägårdh, G. 1989. Membrane cleaning, *Desalination* **71**, 325-335.
- Traina, S.J., Novak, J., Smeck, N.E. 1990. An ultraviolet absorbance method of estimating the percent aromatic carbon content of humic acids. *J. Environ. Qual.* **19** (1), 151-153.
- Westerhoff, P., Chen, W., Esparza, M. 2001. Fluorescence analysis of a standard fulvic acid and tertiary treated wastewater. *J. Environ. Qual.* **30** (6), 2037-2046.
- Wilson, M.A., Ellis, A.V., Lee, G.S.H., Rose, H.R., Lu, X., Young, B.R. 1999. Structure of molecular weight fractions of Bayer humic substances. 1. Low-temperature products. *Ind. Eng. Chem. Res.* **38** (12), 4663-4674.
- Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, M., Koch, R. S. 2004. *PLS-Toolbox Version 3.5 for use with MATLAB™, Manual*. Eigenvector Research, Inc. Manson, WA, USA.
- Wold, S., Esbensen, K., Geladi, P. 1987. Principal components analysis. *Chemo. and Intell. Lab. Sys.* **2**, 37-52.

- Wold, S., Kettaneh-Wold, N., Skagerberg, B., 1989. Non-linear PLS modelling. *Chemo. Intell. Lab. Syst.* **7**, 53-65.
- Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-Regression: A Basic Tool of Chemometrics. *Chemo. Intell. Lab. Syst.* **58**, 109-30.
- Wyatt, P.J. 1993. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272** (1), 1-40.
- Yan, Y., Li, H., Myrick, M.L. 2000. Fluorescence fingerprint of waters: Excitation-emission matrix spectroscopy as a tracking tool. *Appl. Spectrosc.* **54** (10), 1539-1542.
- Yigit, N.O., Civelekoglu, G., Harman, I., Koseoglu, H., Kitis, M. 2009. Effects of various backwash scenarios on membrane fouling in a membrane bioreactor. *Desalination* **237** (1-3), 346-356.
- Yu-Ping, C., Aiken, G., O'Loughlin, E. 1994. Molecular weight, polydispersity, and spectroscopic properties of aquatic humic substances. *Environ. Sci. and Technol.* **28** (11), 1853.
- Yuan, W., Zydney, A. L. 2000. Humic acid fouling during ultrafiltration. *Environ. Sci. and Technol.* **34** (23), 5043-5050.
- Zhao, Y., Li, F., Carvajal, M.T., Harris, M.T. 2009. Interactions between bovine serum albumin and alginate: An evaluation of alginate as protein carrier. *J. Coll. Int. Sci.* **332** (2), 345-353.

- Zheng, X., Ernst, M., Jekel, M. 2009. Identification and quantification of major organic foulants in treated domestic wastewater affecting filterability in dead-end ultrafiltration. *Water Res.* **43** (1), 238-244.
- Zhou, Q., Maurice, P.A., Cabaniss, S.E. 2001. Size fractionation upon adsorption of fulvic acid on goethite: Equilibrium and kinetic studies. *Geochim. Cosmochim. Acta* **65** (5), 803-812.
- Zondervan, E., Betlem, B.H.L., Roffel, B. 2007. Development of a dynamic model for cleaning ultra filtration membranes fouled by surface water. *J. Membr. Sci.* **289** (1-2), 26-31.
- Zularisam, A.W., Ismail, A.F., Salim, M.R., Sakinah, M., Hiroaki, O. 2007. Fabrication, fouling and foulant analyses of asymmetric polysulfone (PSF) ultrafiltration membrane fouled with natural organic matter (NOM) source waters. *J. Membr. Sci.* **299** (1-2), 97-113.

APPENDICES

APPENDIX 1

Rationale for the Selection of PMMA Disposable Cuvettes

The PMMA cuvettes used in this study gradually filter the emission signals captured below the excitation wavelength (Ex): 285 nm. Therefore, the fluorescence intensities at emission wavelength (Em) range: 300 – 600 nm captured below Ex: 285 nm were seen to be lower than emission intensities captured using quartz cuvettes at the same conditions (Figure A - 1, Figure A - 2 and Figure A - 3). In spite of this difference, PMMA cuvettes provided sufficient spectral information necessary to distinguish different fluorescing elements of the NOM in water (Peiris *et al.*, 2008 and 2010). Although this filtering effect of PMMA cuvettes partially removed some spectral information related to AHA and GRW below Ex: 285 nm, it also improved the fluorescence signal quality by filtering the regions of deteriorating signal to noise ratios around Ex: 250 nm (Stedmon *et al.*, 2003). Removal of these regions therefore assisted in minimizing the sensitivity of the fluorescence-based characterization methods developed in this study to measurement noise. Disposable PMMA-based cuvettes or microplates would likely be the material of choice for analysis of large numbers of water samples encountered in practice or that would be used if this approach were to be automated, and hence were used throughout this study.

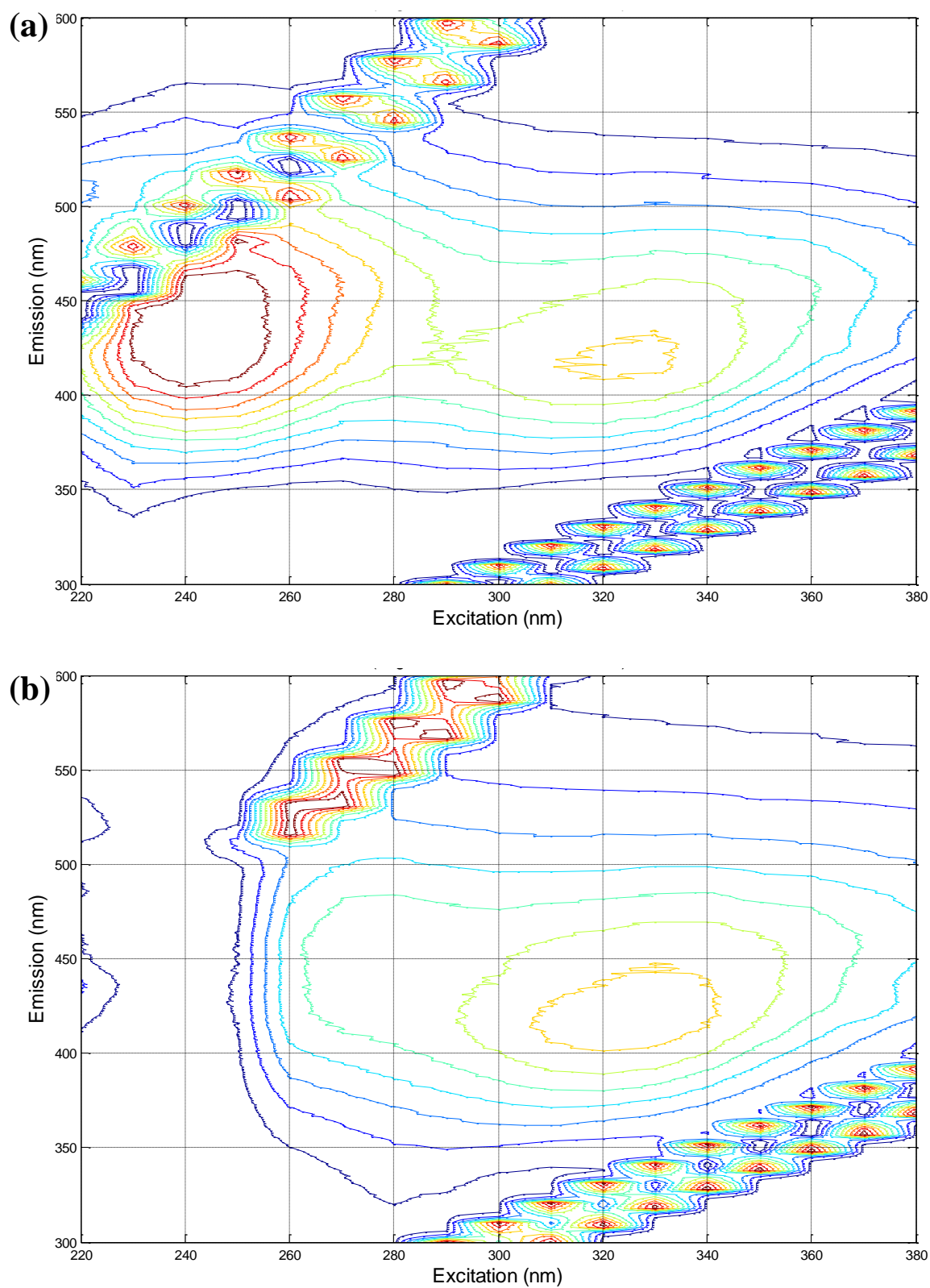


Figure A - 1 Fluorescence EEMs of GRW taken with (a) Quartz cuvettes (above) and (b) PMMA cuvettes (below).

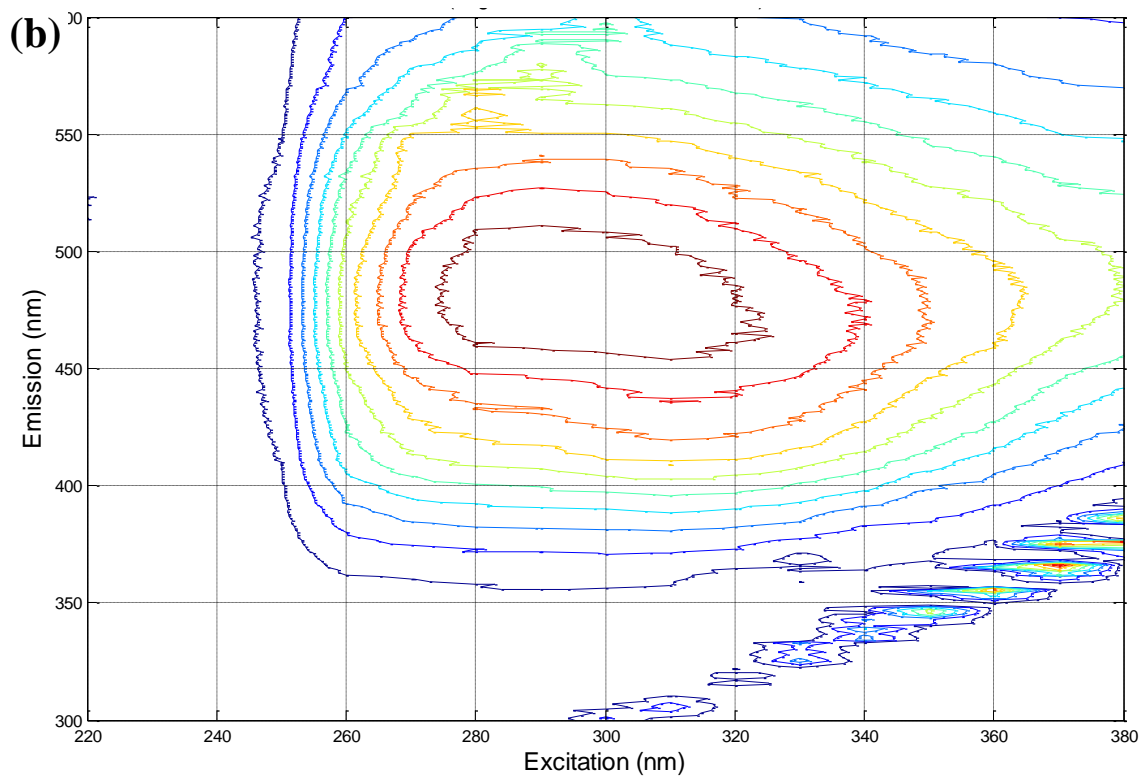
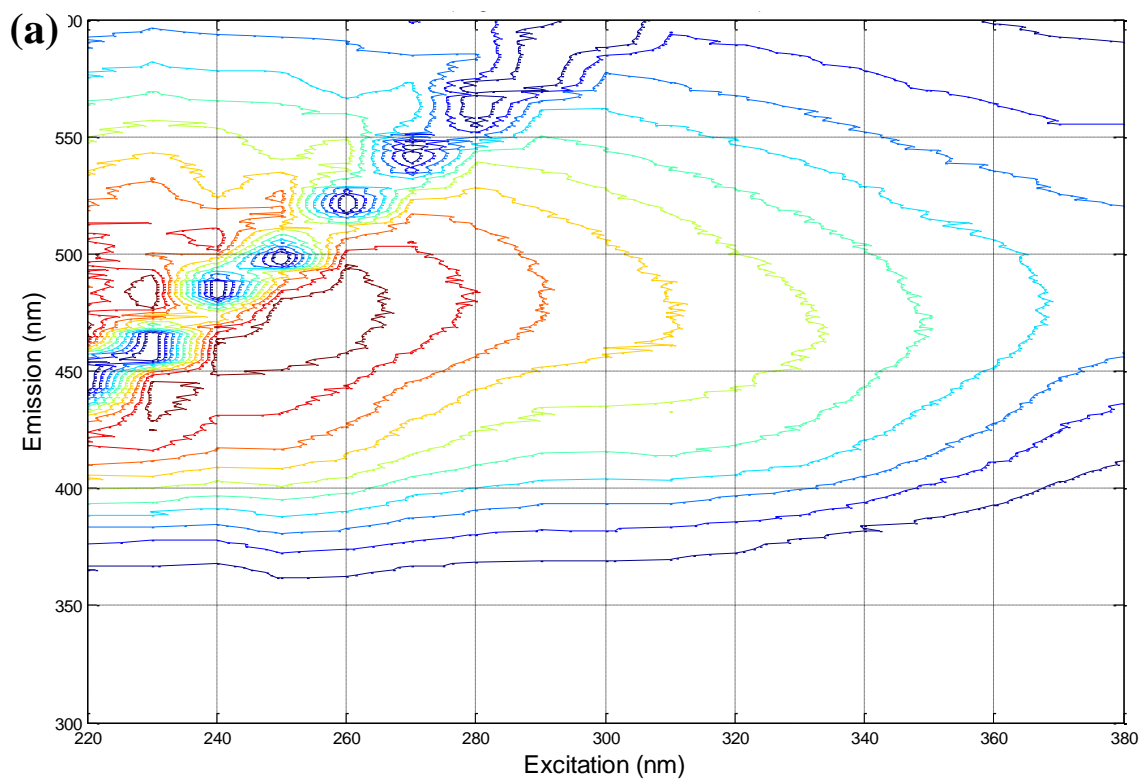


Figure A - 2 Fluorescence EEMs of Aldrich humic acid (4 mg/L) taken with (a) Quartz cuvettes (above) and (b) PMMA cuvettes (below).

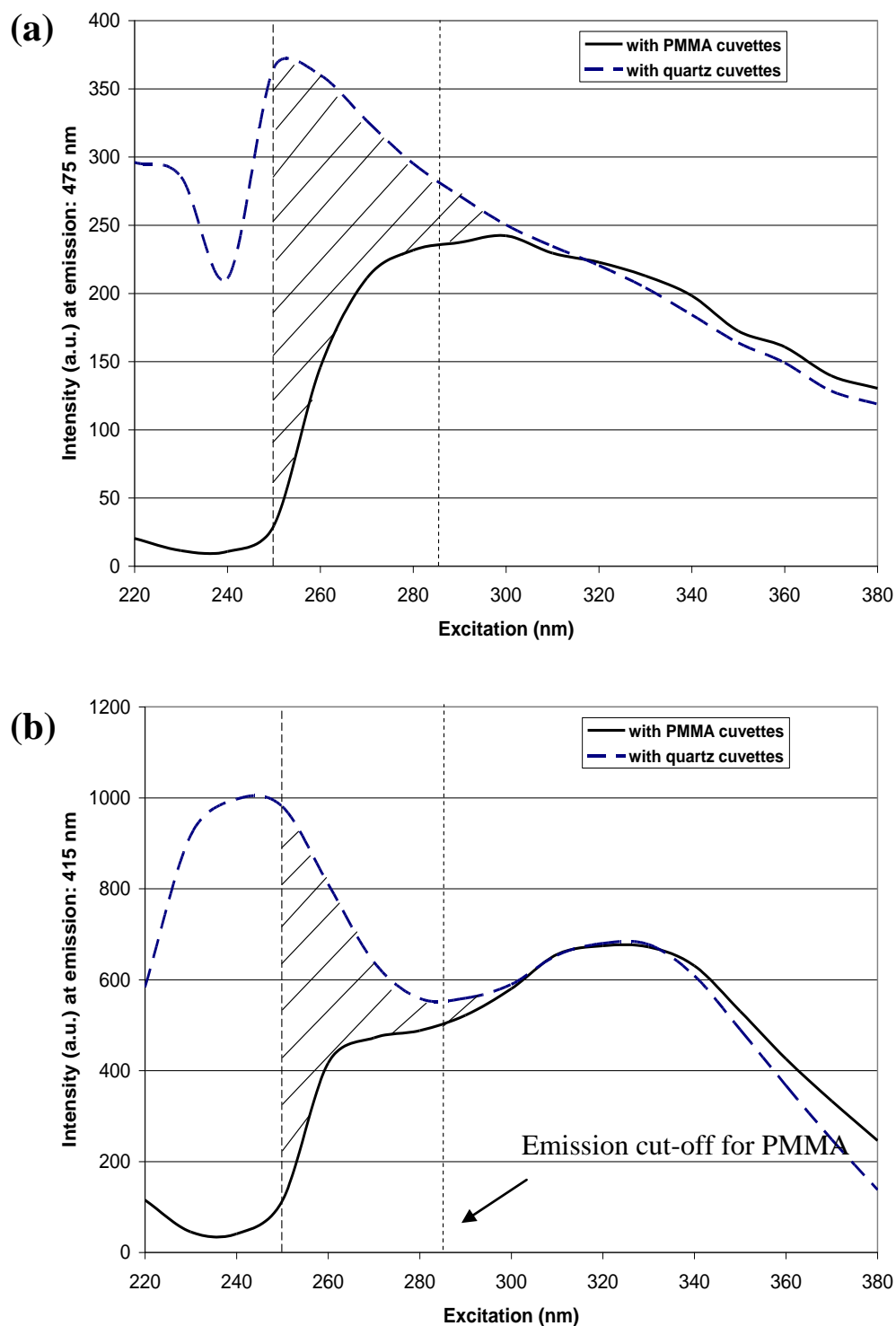


Figure A - 3 Comparison of the fluorescence EEM cross sections of (a) AHA (taken at emission = 475 nm) and (b) GRW (taken at emission = 415 nm) obtained with PMMA and Quartz cuvettes. Hatched-regions denote the fluorescence filtering effect caused by PMMA cuvettes in the excitation wavelength range used in this research.

APPENDIX 2

Selected Photographs from UF Experiments

(a)



(b)



Figure A - 4 Photographs of bench-scale cross flow UF set-up. (a) A view of the UF membrane cell holder including other instrumentation and (b) feed tank connected to the diaphragm pump.

(a)



(b)

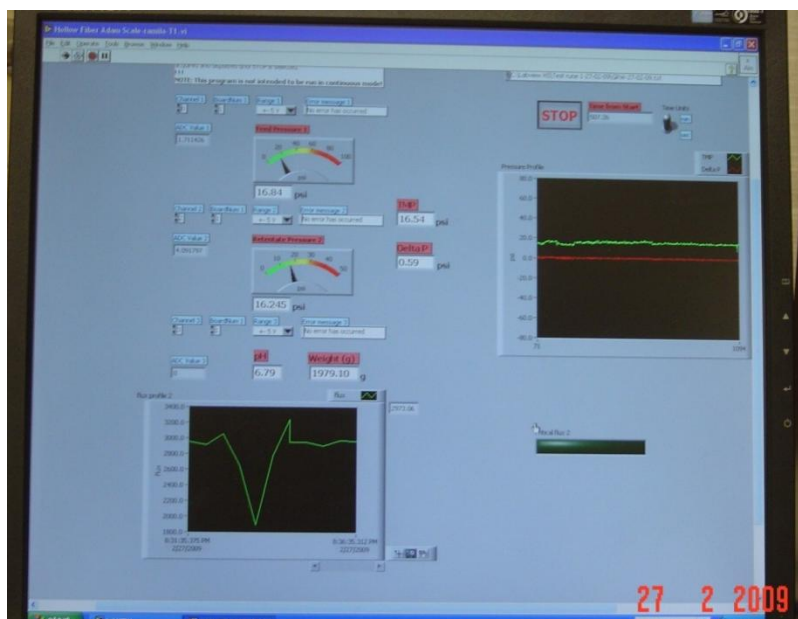
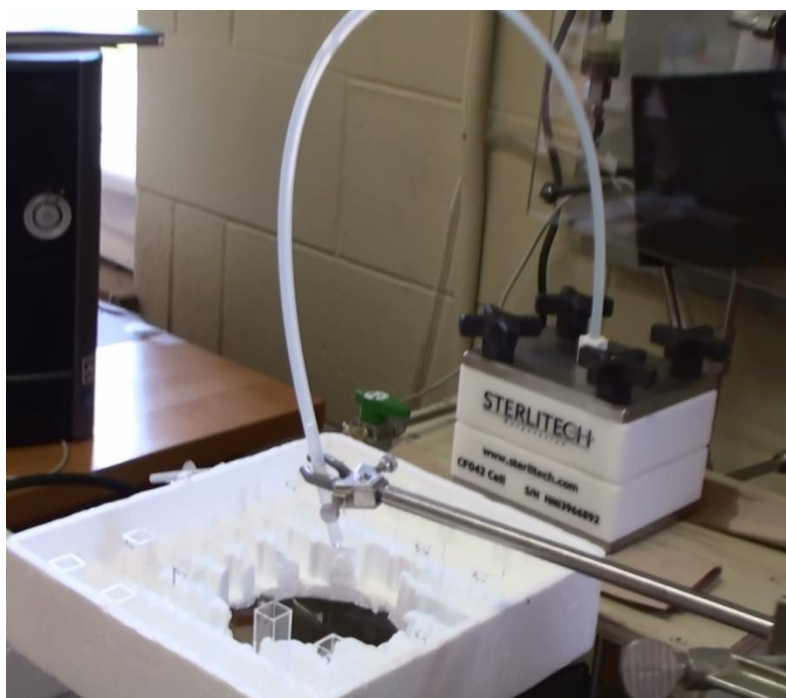


Figure A - 5 (a) A close-up view of the UF instrumentation panel and (b) a computer display of the LabView-based interface used for data logging and monitoring purposes.

(a)



(b)

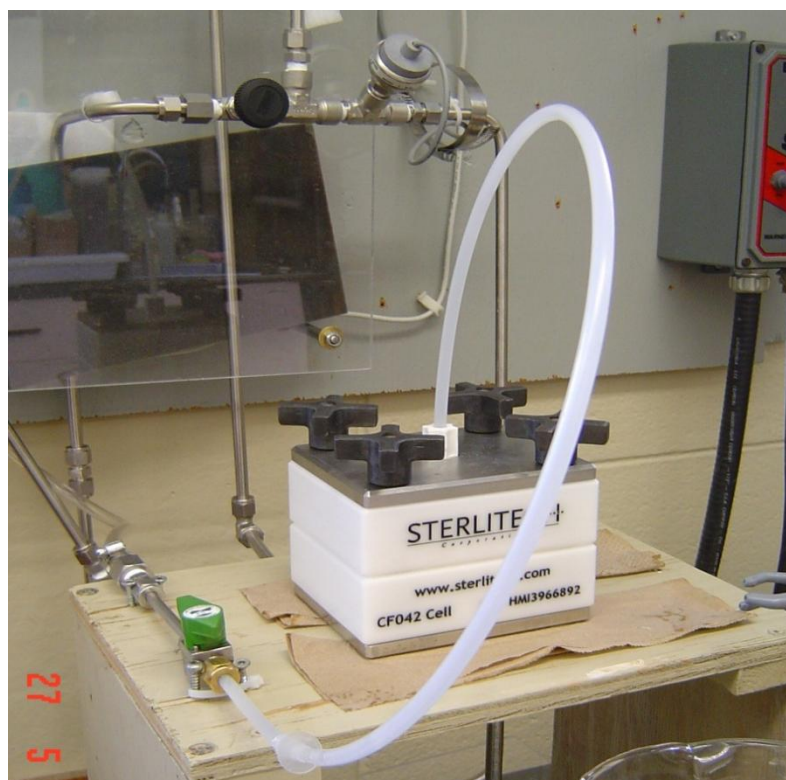


Figure A - 6 Photographs showing (a) sample collection for fluorescence analysis during permeation (b) back-washing of membranes assisted by compressed nitrogen gas.

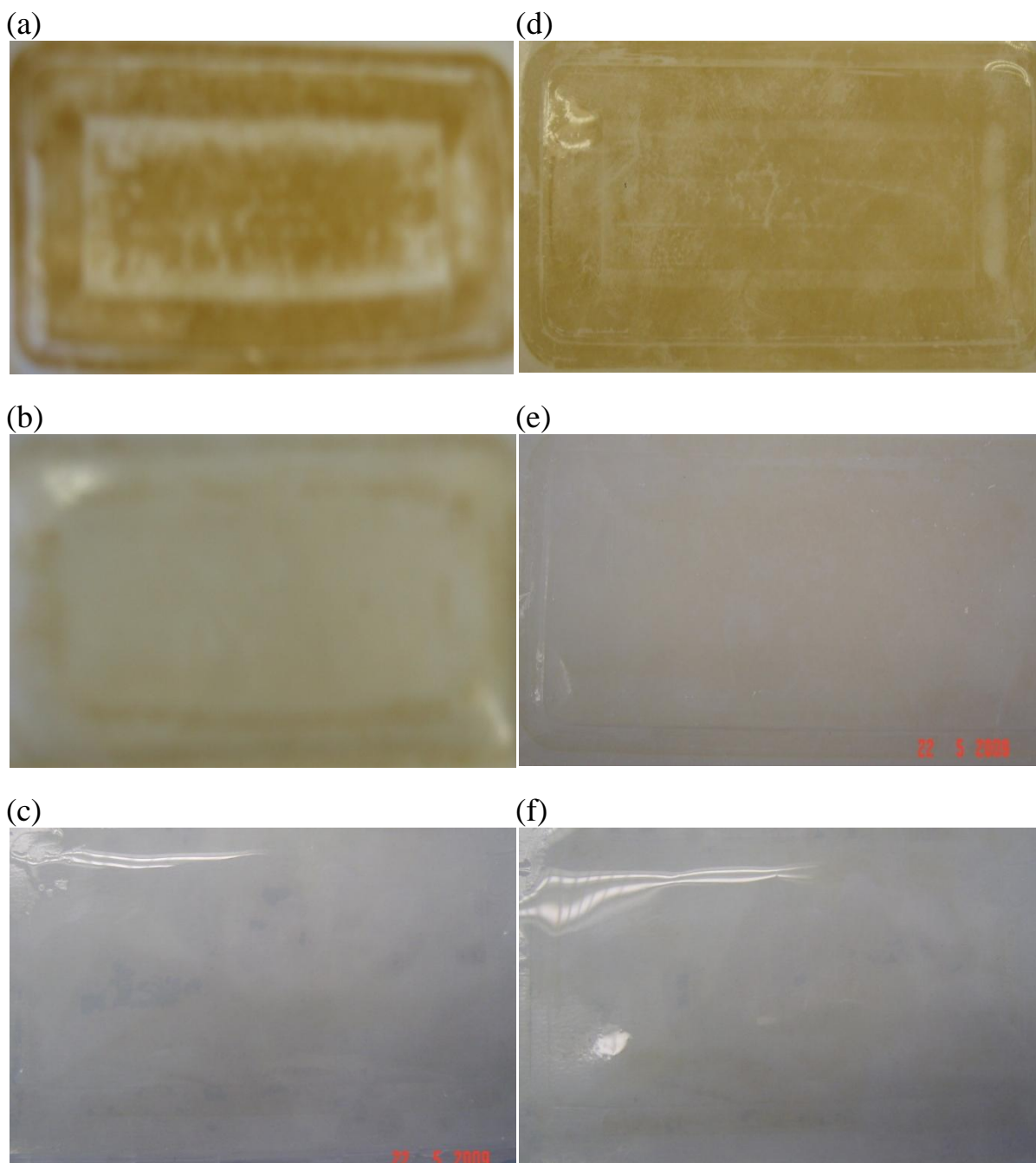


Figure A - 7 Photographs showing membrane surfaces of (a and d) fouled, (b and e) physically cleaned and (c and f) chemically cleaned (i.e. using ethanol) membranes for 60 kDa (left) and 20 kDa (right) UF membranes. These physical and chemical cleaning procedures are explained in Section 9.2.5.

APPENDIX 3

Contact Angle Measurements of 20 kDa and 60 kDa UF Membranes at Different Stages

Table A - 1 Contact angle measurements of 20kDa and 60 kDa UF membranes at different stages.

20 kDa				60 kDa				
Virgin	Fouled (267 min)	Back- washed	EtoH- cleaned	Virgin	EtoH- treated	Fouled (267 min)	Back- washed	EtoH- cleaned
72 ± 2	78 ± 2	74 ± 2	73 ± 2	80 ± 1	75 ± 4	80 ± 1	76 ± 1	75 ± 2

* water used - GRW: 21-05-09; Measurements involved 6 repeated measurements (n=6) at different locations of the active membrane surfaces. EtOH denotes ethanol.

Note: 60 kDa membranes were treated by using a 50 % vol/vol ethanol solution as explained in Section 9.2.2 before filtration as advised by the supplier. According to Table A - 1, 20 KDa (virgin) and 60 kDa (ethanol treated) membranes had similar contact angles. Contact angle measurements of both of these membranes at different stages of the filtration also appeared to be similar in magnitude.

APPENDIX 4

Typical LC-OCD Chromatographs for the Identification of Ultrafiltration Membrane Foulants

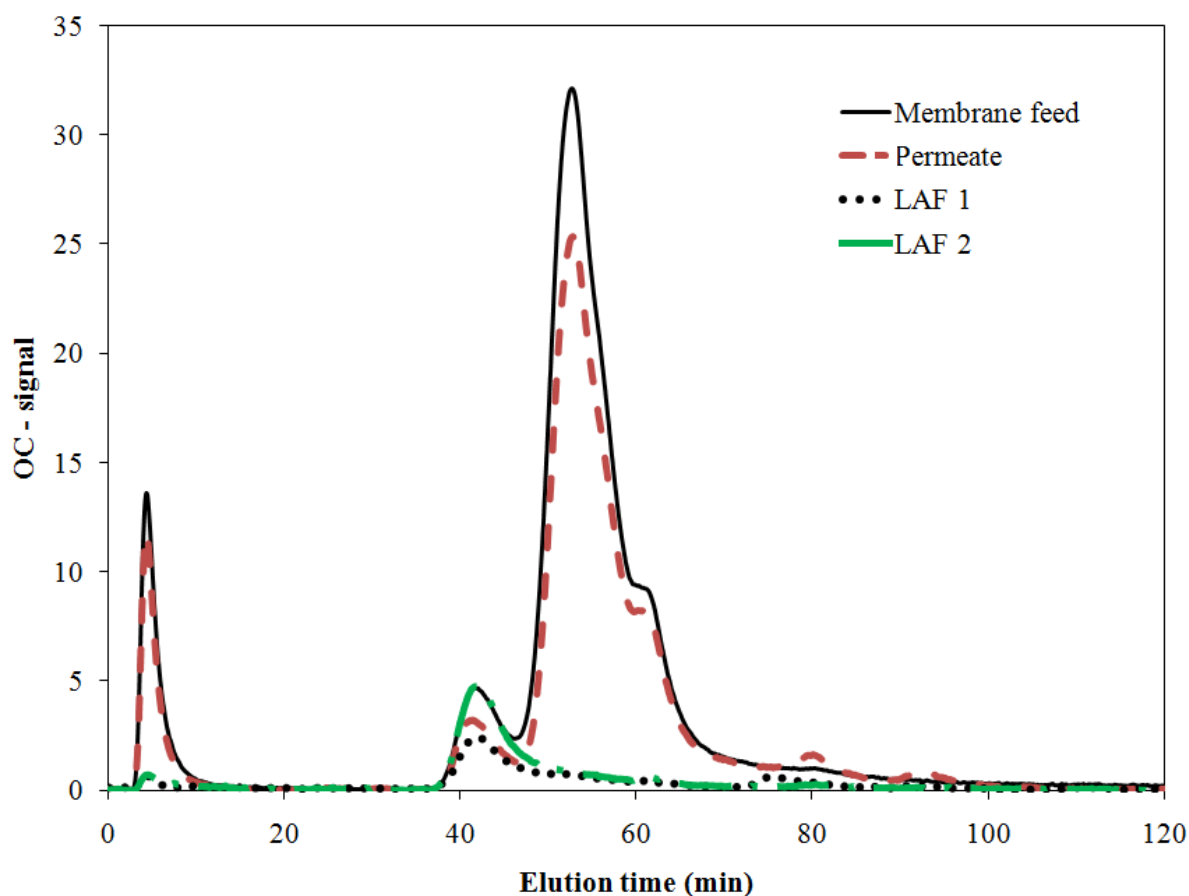


Figure A - 8 LC-OCD chromatographs of feed water (filtered GRW using a 200 micron filter), permeate and loosely attached foulant (LAF 1) material obtained for a 20 kDa UF experiment conducted using GRW obtained on November 23, 2009. LAF 2 was collected from a 60 kDa UF experiment conducted using GRW obtained on October 25, 2009. The procedure followed for the extraction of LAF is explained in Section 9.2.5.